BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

(19) Europäisches Patentamt
Eur pean Patent Office
Office eur péen des brevets

(11) EP 0 781 848 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
02.07.1997 Bulletin 1997/27

(21) Application number: **96120899.8**

(22) Date of filing: **27.12.1996**

(51) Int. Cl.[6]: C12N 15/62, C12N 9/12, C12N 15/48, C07K 14/195, C07K 14/20, C12N 15/54, C07K 14/15, C12N 15/31

(84) Designated Contracting States:
AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE

(30) Priority: **28.12.1995 JP 352225/95**

(71) Applicant: **FUJIREBIO Inc.**
Shinjuku-ku, Tokyo 163-07 (JP)

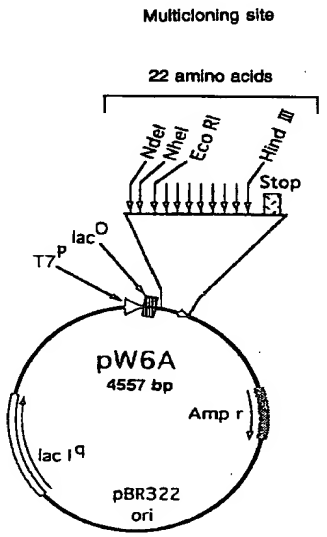
(72) Inventors:
- **Ueno, Eiichi,**
  c/o Fujirebio Inc.
  Shinjuku-ku, Tokyo (JP)
- **Fujii, Nobuyuki,**
  c/o Fujirebio Inc.
  Shinjuku-ku, Tokyo (JP)
- **Okada, Masahisa,**
  c/o Fujirebio Inc.
  Shinjuku-ku, Tokyo (JP)

(74) Representative: **Hansen, Bernd, Dr. Dipl.-Chem. et al**
Hoffmann, Eitle & Partner,
Patentanwälte,
Arabellastrasse 4
81925 München (DE)

(54) Fused DNA sequence, fused protein expressed from said fused DNA sequence and method for expressing said fused protein

(57) Disclosed are a fused DNA sequence which comprises a DNA sequence of a heat-resistant protein, fused directly or indirectly to a DNA sequence coding a selected protein or peptide, a fused protein expressed from the fused DNA sequence, and a method for expressing the fused protein.

Fig. 1



T7 P : T7 promoter
lac O : lac operator
lac I q : lac repressor
↓ : arrows of the multicloning site represent restriction enzyme sites

EP 0 781 848 A2

Description

BACKGROUND OF THE INVENTION

This invention relates to expression of a fused protein, more specifically to a fused DNA sequence including a DNA sequence coding a heat-resistant protein, a fused protein expressed by said fused DNA sequence, and a method for expressing said fused protein.

Progress in genetic engineering has enabled analysis of a protein which has been purified from a natural substance, at a genetic level and artificial amplification of a desired protein (Itakura et al., Science, vol. 198, p. 1056 (1977)). By application of a DNA sequence to which thioredoxin (hereinafter referred to as "TRX" in the specification) (International Provisional Patent Publication No. 507209/1993) or glutathione-S-transferase (hereinafter referred to as "GST" in the specification) (International Provisional Patent Publication No. 503441/1989) which has been invented thereafter is fused, even a protein which is inherently expressed with difficulty can be expressed, and a technique of expressing a fused protein has been used widely.

TRX and GST can be applied to fusion and expression of various proteins which are expressed with difficulty, but even in GST which has been essentially used for the purpose of expressing a soluble fused protein, a fused protein becomes insoluble depending on a protein to be fused so that productivity is lowered, or a fused protein to which TRX is fused may have a problem that a nonspecific reaction is liable to occur. Therefore, it has been desired to provide a fused protein having further excellent operatability and productivity.

SUMMARY OF THE INVENTION

Thus, an object of the present invention is to provide a novel fused DNA sequence having excellent operatability and productivity for expressing a desired protein or peptide, a fused protein expressed from said fused DNA sequence, and a method for expressing the fused protein using said fused DNA sequence.

The present inventors have studied intensively in order to solve the problems in the art and consequently found that when a DNA sequence coding a selected protein or peptide and a DNA sequence coding a heat-resistant protein are fused directly or indirectly and a fused protein is expressed from the resulting fused DNA sequence, the productivity of the desired protein or peptide is raised, and said fused protein has heat resistance to make a purification step simple and easy, to accomplish the present invention.

That is, the present invention relates to a fused DNA sequence comprising a DNA sequence coding a heat-resistant protein or peptide, fused directly or indirectly to a DNA sequence coding a selected protein or peptide, a fused protein expressed by said fused DNA sequence, and a method for expressing the fused protein using said DNA sequence.

The fused protein of the present invention has high solubility and can maintain even heat resistance derived from heat-resistant protein genes. Because of such a characteristic of the fused protein, when the fused protein is purified, unnecessary substances can be removed simply and easily by heat treatment so that the fused protein can be obtained with good yield.

In the case of TRX derived from *Escherichia coli* and GST derived from *Schistosoma japonicum*, which have been widely used as a fused protein, *Escherichia coli* and *Schistosoma japonicum* can live in bodies of mammals and other creatures so that when a fused protein using TRX or GST is used as an antigen of an immunoreaction, a nonspecific reaction due to *Escherichia coli* or *Schistosoma japonicum* might be caused. To the contrary, the great characteristic of the fused protein of the present invention resides in that a heat-resistant protein derived from a thermophilic bacterium which cannot live in living bodies of mammals and other creatures is used so that even when the fused protein of the present invention is used as an antigen of an immuno-reaction, a nonspecific reaction derived from the fused protein is caused with difficulty.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a detailed view of an expression vector pW6A.
Fig. 2 is a detailed view of an expression vector pWF6A.
Fig. 3 is a graph showing the reactivity of a fused protein and a negative specimen.
Fig. 4 is a graph showing the reactivity of a HTLV-I-fused protein and a positive specimen.
Fig. 5 is a graph showing the reactivity of a HTLV-II-fused protein and a positive specimen.
Fig. 6 is a graph showing the reactivity depending on concentration of a HTLV-I-fused protein.
Fig. 7 is a graph showing the reactivity depending on concentration of a HTLV-II-fused protein.
Fig. 8 is a graph showing the activity of a fused protein in a supernatant subjected to heat treatment.
Fig. 9 is a graph showing the activity of a fused protein of precipitates subjected to heat treatment.
Fig. 10 is a view showing the activity of a fused protein after heat treatment and purification.
Fig. 11 is a detailed view of an expression vector pW6AK.

Fig. 12 is a view showing the activity of a fused protein after heat treatment and purification.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

In the following, the present invention is explained in detail.

The DNA sequence coding a heat-resistant protein of the present invention means a DNA sequence coding a protein which is not thermally denatured even at 55 °C or higher, preferably 75 °C or higher. As a specific phenomenon of thermal denaturation, there may be mentioned inactivation or insolubilization of a protein. As the DNA sequence coding a protein which is not thermally denatured at 55 °C or higher, there may be mentioned, for example, a DNA sequence possessed by a thermophilic bacterium which can live at 55 °C or higher. From the properties of an expressed protein and easiness of post-treatment, it is preferred to use a DNA sequence possessed by the so-called highly thermophilic bacterium which can live at 75 °C or higher. As the highly thermophilic bacterium, there may be mentioned, for example, *Thermophilus*, *Sulfolobus*, *Pyrococcus*, *Thermotoga*, *Pyrobaculum*, *Pyrodictium*, *Thermococcus*, *Thermodiscus*, *Metanothermus* and *Metanococcus* (FEMS. MICRO. BIOL. REV., Vol.75, pp.117-124 (1990), ANU. REV. MICROBIOL., Vol.47, pp.627-653 (1993)). As the heat-resistant protein, there may be mentioned, for example, adenyl kinase derived from a *Sulfolobus* bacterium (*Sulfolobus acidocaldalius* Adenylate kinase: Arch. Biochem. Biophys., Vol.207, pp.405-410 (1993)) (hereinafter referred to as "AK" in the specification), DNA polymerase derived from a *Thermophilus* bacterium, ferredoxin derived from a *Pyrococcus* bacterium (*Pyrococcus furiosus* Ferredoxin: Biochemistry, Vol.31, pp.1192-1196 (1992)) (hereinafter referred to as "FDX" in the specification), glucosidase derived from *Pyrococcus furiosus* bacterium (*Pyrococcus furiosus* Glucosidase), rubredoxin derived from *Pyrococcus Furiosus* bacterium (*Pyrococcus furiosus* Rubredoxin: Biochemistry, Vol.30, pp.10885-10895 (1991)), glutamate dehydrogenase derived from *Pyrococcus Furiosus* bacterium (*Pyrococcus furiosus* Glutamate dehydrogenase: Gene, Vol.132, pp.189-197 (1988)), glyceraldehyde phosphate dehydrogenase derived from *Metanothermus fervids* bacterium (*Metanothermus fervids* Glyceraldehyde 3-phosphate dehydrogenase: Gene, Vol.64, p.189-197 (1988)), glutamate synthetase derived from *Metanococcus volate* bacterium (*Metanococcus volate* Glutamate synthetase: Res. Microbiol., Vol.140, pp.355-371 (1989)), L-lactate dehydrogenase derived from *Thermotoga maritina* bacterium (*Thermotoga maritina* L-lactate dehydrogenase: Eur. J. Biochem., Vol.216, pp.709-715 (1993)) and elongation factor derived from *Thermococcus celer* bacterium (*Thermococcus celer* Elongation Factor I-alpha: Nucleic acid res. Vol.18, p.3989 (1990)), but the heat-resistant protein coded by the DNA sequence of the present invention is not limited thereby. DNA coding the heat-resistant protein of the present invention can be purified from these highly thermophilic bacteria, but it can be also synthesized based on a known DNA sequence. For synthesis of DNA of the heat-resistant protein, a known technique such as a β-cyanoethylphosphoamidite method (Sinha et al., Nucleic Acids Bos., Vol.12, p.4539 (1984)) and a method described in Letsinger, R.L. et al., J. Am. Chem. Soc., vol. 88, p. 5319 (1966) may be suitably used. In Examples each of which is an embodiment of the present invention, DNA's of FDX derived from *Pyrocuccus* bacterium and AK derived from *Sulfolobus* bacterium having amino acid sequences shown in SEQ ID NO: 1 and 3, respectively, are synthesized by the β-cyanoethylphosphoamidite method. DNA sequences synthesized are shown in SEQ ID NO: 2 and 4, respectively.

The DNA sequence coding a selected desired protein or peptide of the present invention is not limited to a particular DNA sequence. Any DNA sequence can be used so long as it is a DNA sequence coding a protein or peptide which is desired to be expressed as a fused protein. The present invention is particularly useful when a necessary expression amount of a selected desired protein or peptide can be obtained with difficulty by DNA itself coding said protein or peptide.

The fused DNA sequence of the present invention can be fused by using a known method such as a ligation method and a linker ligation method. When fusion is carried out, the DNA sequence of a selected desired protein or peptide and the DNA sequence of the heat-resistant protein may be fused directly or may be fused indirectly, if necessary. In the case of indirect fusion, a linker sequence is inserted between the DNA sequence coding a desired protein or peptide and the DNA sequence coding the heat-resistant protein. As said linker sequence, there can be used a sequence coding a polypeptide for bonding a desired protein or peptide and the heat-resistant protein to each other and a sequence coding a polypeptide which can be cleaved or digested selectively by a known chemical method or enzymatic method. When the linker sequence is inserted between the DNA sequence coding a desired protein or peptide and the DNA sequence coding the heat-resistant protein, only a selected desired protein or peptide portion can be also purified by, after the fused protein is expressed, cleaving or digesting the linker sequence by using a chemical means such as bromocyan or an enzymatic means such as thrombin or a factor Xa.

In order to express the fused protein of the present invention, a common technique of genetic engineering can be used. For example, the fused DNA sequence of the present invention is inserted into a vector which is suitable for expression, said vector is introduced into a culture host, and expression of the fused protein is induced. After the host is grown by culture or the like, sonication of the host and purification such as a column operation are carried out to obtain a desired fused protein or peptide. Host cells to be used may be any cells such as bacterial cells, eucaryotic cells and mammal cells so long as they are cells which can express a foreign protein or peptide, and there may be mentioned, for example, *Escherichia coli*, yeast, *Bacillus subtilis*, *Baculo* virus and COS cells.

The fused protein of the present invention may be used as such as a fused protein, or a desired protein or peptide portion thereof obtained by separation and purification may be used.

EXAMPLES

The present invention is described in detail by referring to Reference examples and Examples.

Example 1 Preparation of FDX-expressing vector pWF6A

By using 8 primers of 53 mer prepared based on a known DNA sequence of *Pyrococcus furiosus* FDX by using a DNA synthesizer (Model 392, trade name, manufactured by PERKIN ELMER Co.), genes of *Pyrococcus furiosus* FDX were synthesized by the assemble PCR (polymerase chain reaction) method. In the assemble PCR method, a Taq polymerase (produced by Toyobo Co.) was used, and the total base number of 248 bp was amplified under conditions of 30 cycles of 94 °C - 1 minute, 55 °C - 1 minute and 72 °C - 1 minute. A NdeI site was added to 5'-end, a restriction enzyme EcoRI was added to 3'-end, and a thrombin-cut site was added to C terminal. This fragment was integrated into the NdeI and EcoRI sites of 4.6 Kb of a pW6A vector prepared from pGEMEX-1 (trade name, produced by Promega Co.) and pGEX-2T (trade name, produced by Pharmacia Biotec Co.) to prepare pWF6A as a vector expressing FDX. A detailed view of pW6A is shown in Fig. 1, and a detailed view of pWF6A is shown in Fig. 2. pWF6A contains, at the NdeI and EcoRI sites, genes of a fused protein comprising 96 amino acids including 67 amino acids derived from FDX, 10 amino acids derived from a thrombin-cleaved site and 19 amino acids derived from multi cloning site of pW6A. The base sequence of the inserted fragment was confirmed by a DNA sequence kit (trade name: Sequenase kit Ver. 2.0, produced by Amersham United States Biochemical Co.). DNA sequence of the FDX inserted into pW6A and amino acids sequence coded by said sequence are shown in SEQ ID NO: 1 and SEQ ID NO: 2, respectively, and DNA sequence of the pW6A is shown in SEQ ID NO: 5. In the sequence table, ATG of the restriction enzyme site NdeI is shown as 1 and sequences up to the stop codon of a multi-cloning site are shown. The expression "***" in the amino acid sequence means the stop codon. pWF6A was introduced into host *Escherichia coli* and then cultured for 2 hours in a medium (hereinafter referred to as "the LB medium" in the specification) containing 1 % of bactotryptone, 0.5 % of yeast extract, 1 % of sodium chloride and 50 μg/ml of ampicillin and having pH 7.5. Thereafter, 1 mM isopropyl thiogalactopyranoside (hereinafter referred to as "IPTG" in the specification) was added thereto, and the mixture was cultured for 2 hours to induce expression. 10 mM Tris-hydrochloride having pH 7.5 and 1 mM ethylenediaminetetraacetic acid (hereinafter abbreviated to as "EDTA" in the specification) (in the following, this buffer is referred to as "a TE buffer" in the specification) were added to the precipitates of *Escherichia coli*, the precipitates were sonicated, and 15 % sodium dodecylsulfate-polyacrylamide gel electrophoresis (hereinafter referred to as "SDS-PAGE") according to the Laemmli method was carried out. By Coomassie brilliant blue staining (hereinafter referred to as "CBB staining" in the specification), a band was confirmed at about 22 Kda, and FDX of *Pyrococcus furiosus* forming a dimer was recognized.

Example 2 Purification of FDX

pWF6A prepared in Example 1 was introduced into host *Escherichia coli* and then cultured under conditions of using the LB medium at 37 °C. By preculture, a concentration of *Escherichia coli* in a culture broth was made to have such turbidity that absorbance at a wavelength of 600 nm was about 1.0, 1 mM IPTG was added thereto to induce expression. After the mixture was cultured for 3 hours, centrifugation was carried out to recover *Escherichia coli*. 200 ml of a 50 mM Tris-hydrochloride buffer (hereinafter referred to as "the Tris buffer" in the specification) having pH 8.0 was added to recovered *Escherichia coli*, followed by sonication treatment under ice cooling. After centrifugation, the expressed fused protein was recovered in the supernatant as a soluble component. When this supernatant was subjected heat treatment at 85 °C for 15 minutes, about 80 % of the *Escherichia coli* protein was thermally denatured and precipitated, and 90 % or more of FDX was recovered in the centrifugation supernatant after the heat treatment.

This supernatant was purified by ion exchange using a QFF anion exchange column (trade name, manufactured by Pharmacia Biotec Co.) equilibrated with the Tris buffer. When the supernatant was eluted by a column equilibrated buffer containing sodium chloride, FDX was recovered at a concentration of about 0.3 M sodium chloride-eluted fraction. Then, this FDX fraction was purified by using a RESOURCE RPC column (trade name, manufactured by Pharmacia Biotec Co.) equilibrated with 20 mM sodium hydroxide. When the fraction was eluted by acetonitrile, purified FDX was recovered at a concentration of about 10 % acetonitrile-eluted fraction.

Reference example 1 Purification of TRX

pWT8A prepared as a vector expressing TRX in the same manner as in pWF6A prepared in Example 1 was introduced into host *Escherichia coli* and then cultured under conditions of using the LB medium at 37 °C. After the same induction of expression as in Example 1 was carried out, *Escherichia coli* was recovered by centrifugation. An osmotic

shock was given to recovered *Escherichia coli*, and TRX existing at a periplasmic fraction was extracted. Extracted TRX was subjected to first purification by using a RESOURCE RPC column (trade name, manufactured by Pharmacia Biotec Co.) equilibrated with 20 mM sodium hydroxide. When TRX was eluted by acetonitrile, TRX was recovered at a concentration of about 10 % to 20 % acetonitrile-eluted fraction. Recovered TRX was dialyzed to 4 M guanidine hydrochloride and then subjected to second purification by using the reverse phase column under the same conditions. Similarly as in the first purification, purified TRX was recovered at a concentration of about 10 % to 20 % acetonitrile-eluted fraction.

Example 3 Specificity test of FDX and TRX by the western blotting method

An anti-*Escherichia coli* antibody was supposed as a non-specific reaction substance, and the reactivities of FDX purified in Example 2 and TRX purified in Reference example 1 were examined.

A SDS-solubilized material of *Escherichia coli* DH5α, a supernatant of *Escherichia coli* DH5α sonicated and a SDS-solubilized material of *Escherichia coli* to which a pW50 vector (made by Fuji Rebio) was introduced were used as immunogen and immunized to 3 rabbits to prepare the total 9 kinds of the respective anti-*Escherichia coli* rabbit serums. FDX purified in Example 2 and TRX purified in Reference example 1 were subjected to SDS-PAGE according to the Laemmli method and then transferred to nitrocellulose membranes. After blocking the protein portion adsorbed to the nitrocellulose membranes with 1 % skim milk dissolved in PBS, the western blotting method was carried out by using the above 9 kinds of the anti-*Escherichia coli* rabbit serums diluted 500 times, respectively, as primary antibodies, and using a peroxidase (hereinafter referred to as "POD" in the specification)-labeled anti-rabbit antibody as a secondary antibody. For coloring, 4-chloro-1-naphthol and hydrogen peroxide were used. At the portion corresponding to the molecular weight of FDX, no substance reacting with the anti-*Escherichia coli* rabbit antibody was confirmed, but at the portion corresponding to the molecular weight of TRX, among 9 kinds of the anti-*Escherichia coli* rabbit serums, 6 kinds of the serums in which the supernatant of *Escherichia coli* DH5α sonicated and the SDS-solubilized material of *Escherichia coli* into which the pW50 vector was introduced were used as immunogen were reacted, respectively.

In the same manner as described above, the western blotting method was carried out by 25 samples of human specimen HTLV-I/II mix panel 204 serums (trade name, produced by Boston Biomedica Co.) diluted 50 times, respectively, as primary antibodies, and using POD-labelled anti-human IgG as a secondary antibody. Reactivities at sites where FDX was transferred was not confirmed, but the reactions of 2 samples among 25 samples at sites where TRX was transferred were confirmed. The results are shown in Table 1.

Table 1

Specimen No.	Intensity of reaction (+, -) by western blotting	
	FDX	TRX
PRP-204-01	-	-
PRP-204-02	-	-
PRP-204-03	-	-
PRP-204-04	-	-
PRP-204-05	-	-
PRP-204-06	-	-
PRP-204-07	-	-
PRP-204-08	-	-
PRP-204-09	-	-
PRP-204-10	-	-
PRP-204-11	-	-
PRP-204-12	-	+
PRP-204-13	-	-
PRP-204-14	-	-
PRP-204-15	-	-
PRP-204-16	-	-
PRP-204-17	-	-
PRP-204-18	-	-
PRP-204-19	-	-
PRP-204-20	-	-
PRP-204-21	-	-
PRP-204-22	-	-
PRP-204-23	-	+
PRP-204-24	-	-
PRP-204-25	-	-
+: positive, -: negative		

Example 4 Specificity test of FDX and TRX by the ELISA method using human specimens

On ELISA plates (produced by Becton Deckinson Co.) were sensitized each 50 µl of 25 µg/ml of FDX purified in Example 2 and TRX purified in Reference example 1, respectively.

After blocking the protein portion adsorbed onto wells of the ELISA plate with 1 % skim milk, a specificity test according to the ELISA method was carried out by using the human specimens produced by Boston Biomedica Co. diluted 500 times used in Example 3 as primary antibodies and POD-labelled anti-human IgG as a secondary antibody. For coloring, ABTS and hydrogen peroxide were used. The measurement results were shown by difference between absorbances at a wavelength of 405 nm and a wavelength of 492 nm (difference between absorbances was described as A405/492 nm). In the reactions with the specimens, whereas there was no specimen exceeding twice of a blank in

the case of FDX, the specimens exceeding twice of a blank were confirmed in 6 samples among 25 samples in the case of TRX. FDX derived from *Pyrococcus furiosus* was different from TRX derived from *Escherichia coli* in that neither nonspecific reaction nor cross reaction derived from *Escherichia coli* was recognized. The results are shown in Fig. 3.

Example 5 Expression of FDX-fused HTLV-I p19-fused protein and FDX-fused HTLV-II p19-fused protein

From infected cell lines expressing HTLV-I and HTLV-II, genomic DNA was extracted by the method of Molecular Cloning by J. Sambrook et al. Next, by using a primer to which EcoRI and BamHI sites were added, the PCR method was carried out in the same manner as in Example 1 to obtain about 400 bp of p19DNA fragments in the respective gag regions. These fragments were integrated into pWF6A to prepare pWFIP19 as a vector expressing p19 of HTLV-I and pWFIIP19 as a vector expressing p19 of HTLV-II. DNA sequences of the FDX-fused HTLV-I p19 and FDX-fused HTLV-II p19 each of which is inserted into the vectors are shown in SEQ ID NO: 6 and 8, respectively, and amino acids sequences coded by said DNA sequences are shown in SEQ ID NO: 7 and 9, respectively. In the same manner as in Example 1, these vectors were introduced into *Escherichia coli*, and expression of the respective fused proteins was induced. Samples for electrophoresis were prepared under the same conditions as in Example 1. After subjecting to 12.5 % SDS-PAGE according to the Laemmli method, one sheet of gel was subjected to CBB staining, and the other sheet was transferred to nitrocellulose membranes by the method shown in Example 3. By using an anti-native HTLV-I p19 monoclonal antibody (a GIN-7 antibody, Tanaka, Y. et al., Gann., Vol.74, pp.327 to 330 (1983)) or an anti-native HTLV-II p19 monoclonal antibody as a primary antibody, and a POD-labeled antimouse IgG as a secondary antibody, these were reacted with the fused proteins by the same method as in Example 3 and coloring was carried out by using 4-chloro-1-naphthol and hydrogen peroxide, expression of the fused proteins reacting with the respective monoclonal antibodies corresponding to the respective fused proteins was recognized. These fused proteins gave a band at about 34 Kda which was the same position as that of the CBB-stained gels. The expression amounts of the FDX-fused HTLV-I p19 antigen and the FDX-fused HTLV-II p19 antigen were increased by several hundreds times as compared with the case where the p19 antigen of HTLV-I and the p19 antigen of HTLV-II were expressed directly.

Example 6 Expression of FDX-fused HTLV-I p20E(gp21)-fused protein and HTLV-II p20E(gp21)-fused protein

By the same method as in Example 5, by using DNA of cells infected with HTLV-I and HTLV-II, about 500 bp of p20E(gp21) DNA fragments in the respective env regions were obtained by the PCR method. These DNA fragments were integrated into EcoRI and BamHI of pWF6A prepared in Example 1 to prepare pWFIE1 as a vector expressing p20E of HTLV-I and pWFIIE10 as a vector expressing p20E of HTLV-II. DNA sequences of the FDX-fused HTLV-I p20E and FDX-fused HTLV-II p20E each of which is inserted into the vectors are shown in SEQ ID NO: 10 and 12, respectively, and amino acids sequences coded by said DNA sequences are shown in SEQ ID NO: 11 and 13, respectively. These vectors were introduced into *Escherichia coli*, and expression of a FDX-fused HTLV-I p20E-fused protein (hereinafter referred to as "FDX-20(I)" in the specification) and a FDX-fused HTLV-II p20E-fused protein (hereinafter referred to as "FDX-20(II)" in the specification) was induced under the same conditions as in Example 1. In the same manner as in Example 1, *Escherichia coli* was sonicated. After subjecting to 12.5 % SDS-PAGE according to the Laemmli method, one sheet of gel was subjected to CBB staining, and the other sheet of gel was transferred to nitrocellulose membranes at 120 mA for 3 hours. After blocking the protein portion adsorbed to the nitrocellulose membranes with a phosphate buffer containing 1 % of BSA (bovine serum albumin), 1 μg/ml of an anti-p20E(gp21) monoclonal antibody (F-10, Sugamura, K. et al., J. Immunol., Vol.132, pp.3180 to 3184 (1984)) reacting with p20E(gp21) antigens of native HTLV-I and HTLV-II was reacted with the fused proteins at room temperature for 1 hour, and then reacted with a POD-labeled anti-mouse IgG at room temperature for 1 hour. Subsequently, when coloring was carried out by using 4-chloro-1-naphthol and hydrogen peroxide, expression of fused proteins reacting with the anti-p20E(gp21) monoclonal antibody corresponding to the respective fused proteins was recognized. These fused proteins gave a band at about 32 Kda which was the same position as that of the CBB-stained gels.

The expression amounts of FDX-20(I) and FDX-20(II) were increased by several hundreds times as compared with the case where p20E of HTLV-I and p20E of HTLV-II were expressed directly.

Example 7 Purification of FDX-20(I)- and FDX-20(II)-fused proteins

pWFIE1 and pWFIIE10 prepared in Example 6 were introduced into host *Escherichia coli*, respectively, and then cultured under conditions of using the LB medium at 37 °C. By preculture, a concentration of *Escherichia coli* in culture broths was made to have such turbidity that absorbance at a wavelength of 600 nm was about 1.0, 1 mM IPTG was added thereto to induce expression. Three hours after IPTG was added, centrifugation was carried out to recover *Escherichia coli*. 200 ml of a 50 mM Tris-hydrochloride buffer containing 1 % Triton X 100 (trade name, produced by Rohm & Haas Co.) and 2 M urea with pH 8.0 was added to recovered *Escherichia coli*, followed by sonication treatment under ice cooling. Centrifugation was carried out to recover insoluble materials (inclusion bodies). The inclusion bodies

were solubilized by using a 4 M guanidine hydrochloride-10 mM dithiothreitol (hereinafter referred to as "DTT" in the specification) solution. The solubilized bodies were purified by a RESOURCE RPC column (trade name, manufactured by Pharmacia Biotec Co.) equilibrated with 20 % acetonitrile and 20 mM sodium hydroxide. When the bodies were eluted by acetonitrile, purified FDX-20(I)- and FDX-20(II)-fused proteins were recovered at a concentration of about 30 to 40 % acetonitrile-eluted fractions, respectively.

Reference example 2 Purification of TRX-fused HTLV-I p20E-fused protein and TRX-fused HTLV-II p20E-fused protein

In the same manner as in Example 6, p20E(gp21) in an env region of HTLV-I or HTLV-II was introduced into the TRX-expressing vector pWT8A prepared in Reference example 1 to prepare pWTIE1 and pWTIIE10, followed by expression. In the same manner as in Example 7, by the purification method using a RESOURCE RPC column (trade name, manufactured by Pharmacia Biotec Co.), a TRX-fused HTLV-I p20E-fused protein (hereinafter referred to as "TRX-20(I)" in the specification) and a TRX-fused HTLV-II p20E-fused protein (hereinafter referred to as "TRX-20(II)" in the specification) were purified.

Example 8 Reactivity test of fused proteins

(1) Test by the western blotting method

By using FDX-20(I) and FDX-20(II) purified in Example 7 and TRX-20(I) and TRX-20(II) purified in Reference example 2, reactivities with human HTLV specimens in the western blotting method were compared.

In the same manner as in Example 3, the western blotting method was carried out by using the human specimen HTLV-I/II mix panel produced by Boston Biomedica Co. diluted 50 times as primary antibodies and POD-labelled human IgG as a secondary antibody. FDX-20(I) and FDX-20(II), and TRX-20(I) and TRX-20(II) were reacted with the same specimens, respectively. The results are shown in Table 2.

Table 2

Specimen No.	Intensity of reaction (+, -) by western blotting			
	FDX-20(I)	TRX-20(I)	FDX-20(II)	TRX-20(II)
PRP-204-01	+	+	+	+
PRP-204-02	-	-	-	-
PRP-204-03	+	+	+	+
PRP-204-04	-	-	+	+
PRP-204-05	+	+	-	-
PRP-204-06	-	-	-	+
PRP-204-07	+	+	+	+
PRP-204-08	-	-	-	-
PRP-204-09	+	+	-	-
PRP-204-10	+	+	+	+
PRP-204-11	+	+	+	+
PRP-204-12	++	++	++	++
PRP-204-13	+	+	+	+
PRP-204-14	-	-	+	+
PRP-204-15	+	+	+	+
PRP-204-16	-	-	+	+
PRP-204-17	+	+	+	+
PRP-204-18	+	+	+	+
PRP-204-19	+	+	-	-
PRP-204-20	-	-	-	-
PRP-204-21	+	+	+	+
PRP-204-22	+	+	+	+
PRP-204-23	+	+	+	+
PRP-204-24	+	+	+	+
PRP-204-25	+	+	+	+

+: positive,
++: strongly positive,
-: negative

(2) Comparison by the ELISA method

On ELISA plates (produced by Becton Deckinson Co.) were sensitized each 50 µl of FDX-20(I) and FDX-20(II) purified in Example 7 and TRX-20(I) and TRX-20(II) purified in Reference example 2 at a concentration of 3 µg/ml, respectively.

The ELISA method was carried out by using these ELISA plates and using the human specimens produced by Boston Biomedica Co. diluted 500 times as primary antibodies and POD-labelled anti-human IgG as a secondary antibody in the same manner as in Example 4. FDX-20(I) and FDX-20(II), and TRX-20(I) and TRX-20(II) were reacted with the same specimens. The results are shown in Fig. 4 and Fig. 5.

(3) Test of dependency on concentration by the ELISA method

In order to examine reactivities to the anti-p20E(gp21) monoclonal antibody and a negative serum, 10 µg/ml to 1/2 dilution series of FDX-20(I) and FDX-20(II) purified in Example 7 and TRX-20(I) and TRX-20(II) purified in Reference example 2 were prepared, respectively, and ELISA plates (produced by Becton Deckinson Co.) were sensitized with each 50 µl thereof.

The ELISA method was carried out by using these ELISA plates and using the anti-p20E(gp21) monoclonal antibody diluted 500 times as a primary antibody and POD-labelled anti-mouse IgG as a secondary antibody. With respect to a negative serum, the ELISA method was carried out in the same manner as in Example 4. There was no difference in reactivity to the monoclonal antibody, and the FDX-fused proteins in both cases of HTLV-I and HTLV-II had lower reactivities to the negative serum. The results are shown in Fig. 6 and Fig. 7.

Reference example 3 Preparation of protein in which GST and *Treponema pallidum* 15Kda antigen are fused

From syphilis bacteria (Nichols strain from *Treponema pallidum*) purified from syphilis bacteria-subcultured rabbit testicles, genomic DNA was extracted. By using the extracted DNA as a template, a primer was produced based on the known DNA sequences by using a DNA synthesizer (Model 392, trade name, produced by PERKIN ELMER Co.). By using the primer, about 370 bp of a DNA fragment coding a surface antigen of 15 Kda (hereinafter referred to as "Tp15" in the specification) of *Treponema pallidum* (hereinafter referred to as "Tp" in the specification) was amplified with a thermal cycler (Model PJ1000, trade name, produced by PERKIN ELMER Co.). This DNA fragment was integrated into an EcoRI site of a GST-expressing type vector pWG6A in which DNA sequence of GST had been inserted into pW6A to obtain a vector pWGTp15 expressing a protein in which GST and Tp15 were fused (hereinafter referred to as "GST-15" in the specification). DNA sequence of the GST-15 inserted into the vector is shown in SEQ ID NO: 14 and amino acids sequence coded by said DNA sequence is shown in SEQ ID NO: 15. In the same manner as in Example 1, the vector was introduced into *Escherichia coli*, and expression of GST-15 was induced. A sample for electrophoresis was prepared under the same conditions as in Example 1. After subjecting to 12.5 % SDS-PAGE according to the Laemmli method, one sheet of gel was subjected to CBB staining, and the other sheet was transferred to a nitrocellulose membrane by the method shown in Example 3. By using an anti-Tp15 monoclonal antibody as a primary antibody and a POD-labeled mouse IgG as a secondary antibody, these were reacted in the same method as in Example 3 and coloring was carried out by using 4-chloro-1-naphthol and hydrogen peroxide, a band was given at about 42 Kda which was the same position as that of the CBB-stained gel.

Reference example 4 Preparation of protein in which TRX and Tp15 are fused

A DNA fragment of Tp15 amplified in Reference example 3 was integrated into an EcoRI site of the TRX-expressing type vector pWT8A in which DNA sequence of TRX had been inserted into pW6A to obtain a vector pWTTp15 expressing a protein in which TRX and Tp15 were fused (hereinafter referred to as "TRX-15" in the specification). DNA sequence of the TRX-15 inserted into the vector is shown in SEQ ID NO: 16 and amino acids sequence coded by said DNA sequence is shown in SEQ ID NO: 17. In the same manner as in Example 1, the vector was introduced into *Escherichia coli*, and expression of TRX-15 was induced. A sample for electrophoresis was prepared under the same conditions as in Example 1. After subjecting to 12.5 % SDS-PAGE according to the Laemmli method, one sheet of gel was subjected to CBB staining, and the other sheet was transferred to a nitrocellulose membrane by the method shown in Example 3. By using an anti-Tp15 monoclonal antibody as a primary antibody and a POD-labeled mouse IgG as a secondary antibody, these were reacted in the same method as in Example 3 and coloring was carried out by using 4-chloro-1-naphthol and hydrogen peroxide, a band was given at about 27 Kda which was the same position as that of the CBB-stained gel.

Example 9 Preparation of protein in which FDX and Tp15 are fused

A DNA fragment of Tp15 amplified in Reference example 3 was integrated into an EcoRI, BamHI site of the FDX-expressing type vector pWF6A prepared in Example 1 to obtain a vector pWFTp15 expressing a protein in which FDX and Tp15 were fused (hereinafter referred to as "FDX-15" in the specification). DNA sequence of the FDX-15 inserted into the vector is shown in SEQ ID NO: 18 and amino acids sequence coded by said DNA sequence is shown in SEQ ID NO: 19. In the same manner as in Example 1, the vector was introduced into *Escherichia coli*, and expression of FDX-15 was induced. A sample for electrophoresis was prepared under the same conditions as in Example 1. After subjecting to 12.5 % SDS-PAGE according to the Laemmli method, one sheet of gel was subjected to CBB staining, and the other sheet was transferred to a nitrocellulose membrane by the method shown in Example 3. By using an anti-Tp15 monoclonal antibody as a primary antibody and a POD-labeled mouse IgG as a secondary antibody, these were reacted in the same method as in Example 3 and coloring was carried out by using 4-chloro-1-naphthol and hydrogen

peroxide, a band was given at about 30 Kda which was the same position as that of the CBB-stained gel.

Example 10 Heat resistance test of FDX-15, GST-15 and TRX-15

The vectors expressing FDX-15, GST-15 and TRX-15 prepared in Example 9, Reference example 3 and Reference example 4 were introduced into host *Escherichia coli* and then cultured under conditions of using 1 liter of the LB medium at 37 °C, respectively. By preculture, a concentration of *Escherichia coli* in culture broths was made to have such turbidity that absorbance at a wavelength of 600 nm was about 1.0, 1 mM IPTG was added thereto to induce expression. After the cells were recovered by centrifugation, 200 ml of the Tris buffer was added to the cells. After sonication treatment under ice cooling, fused proteins were recovered in the centrifugation supernatants, respectively. 800 µl of these proteins were taken, respectively, and shaken for 13 minutes in water bath at 40 °C, 50 °C, 60 °C, 70 °C and 80 °C. The respective samples were centrifuged and then separated into supernatants and precipitates, and analysis was carried out by SDS-PAGE and the western blotting method. As a blocking agent of the western blotting method, 1 % skim milk dissolved in PBS was used, and as a primary antibody, an anti-TP rabbit antibody was used. As a secondary antibody, a POD-labelled anti-rabbit antibody was used, and as a coloring agent, 4-chloro-1-naphthol and hydrogen peroxide were used. The result of coloring of western blotting was confirmed by a densitometer. The results are shown in Fig. 8 and Fig. 9. Precipitates of TRX-15 and GST-15 were partially generated at 40 °C by thermal denaturation, about 80 % of TRX-15 and GST-15 were precipitated at 60 °C, and about 100 % of them were precipitated at 70 °C. Almost no precipitate by thermal denaturation of FDX-15 was generated at 40 °C to 80 °C, and even at 80 °C, about 100 % of FDX-15 existed in the supernatant.

Example 11 Purification of FDX-15 by heat treatment

pWFTp15 prepared in Example 9 was introduced into host *Escherichia coli* and then cultured under conditions of using 1 liter of the LB medium at 37 °C. By preculture, a concentration of *Escherichia coli* in culture broths was made to have such turbidity that absorbance at a wavelength of 600 nm was about 1.0, 1 mM IPTG was added thereto to induce expression. The cells were recovered by centrifugation. 200 ml of the Tris buffer was added to the cells, and the cells were sonicated to recover FDX-15 in the centrifugation supernatant. Then, by using a hot plate and a water bath, heat treatment at 70 °C for 10 minutes was carried out to recover FDX-15 in the centrifugation supernatant. The supernatant subjected to heat treatment was purified by a QFF anion exchange column (trade name, manufactured by Pharmacia Biotec Co.) equilibrated with the Tris buffer. When the supernatant was eluted by a column equilibrated buffer containing sodium chloride, FDX-15 was recovered at a concentration of about 0.3 M to 0.4 M sodium chloride-eluted fraction. Then, 10 mM DTT was added to the QFF recovered fraction, and the mixture was purified by using a RESOURCE RPC column (trade name, manufactured by Pharmacia Biotec Co.) equilibrated with a 20 mM sodium hydroxide solution. When the mixture was eluted by acetonitrile, FDX-15 was recovered at a concentration of about 20 % to 25 % acetonitrile-eluted fraction. This reverse phase recovered fraction was concentrated by Centriprep (trade name, manufactured by Amicon Inc.), and the concentrate was subjected to gel filtration by a Superdex 200 column (trade name, manufactured by Pharmacia Biotec Co.). When the filtrate was eluted by a buffer containing 6 M urea, 0.5 M sodium chloride and 20 mM Tris-hydrochloride having pH 8.0, purified FDX-15 was recovered at a molecular weight of about 50,000. By heat treatment at 60 °C, about 80 % of the *Escherichia coli* protein was precipitated by thermal denaturation, but even at 70 °C, almost 100 % of FDX-15 was recovered in the supernatant, and the purification degree was raised by about 5 times only by heat treatment.

Further, GST-15 obtained by introducing pWGTp15 prepared in Reference example 3 into host *Escherichia coli*, carrying out induction and expression operations in the same manner therein and carrying out purification by a common column operation without carrying out heat treatment and FDX-15 purified by heat treatment were subjected to the western blotting method in the same manner as in Example 10 by using an anti-Tp rabbit antibody. It was shown that even though purification by heat treatment was carried out, FDX-15 retained reactivity. The results are shown in Fig. 10.

Example 12 Preparation of AK-expressing vector pW6AK

By using 16 primers of 53 mer prepared based on a known DNA sequence of AK derived from a *Sulfolobus* bacterium by using a DNA synthesizer (manufactured by Perkin Elmer Co.), genes of *Sulfolobus acidocaldarius* AK were synthesized by the assemble PCR method. In the assemble PCR method, a Taq polymerase (produced by Toyobo Co.) was used, and the total base number of 630 bp was amplified under conditions of 30 cycles of 94 °C - 1 minute, 55 °C - 1 minute and 72 °C - 1 minute. A NdeI site was added to 5'-end, a restriction enzyme EcoRI was added to 3'-end, and a thrombin-cut site was added to C terminal. This fragment was integrated into the NdeI and EcoRI sites of 4.6 Kb of a pW6A vector prepared from pGEMEX-1 (trade name, produced by Promega Co.) and pGEX-2T (trade name, produced by Pharmacia Biotec Co.) to prepare pW6AK as a vector expressing AK. A detailed view of pW6AK is shown in Fig. 11. pW6AK contains genes of a fused protein comprising 223 amino acids including 194 amino acids derived from AK, 10

amino acids derived from a thrombin-cleaved site and 19 amino acids derived from multi cloning site of pW6A, at the NdeI and EcoRI sites. The base sequence of the inserted fragment was confirmed by a DNA sequence kit (trade name: Sequenase kit Ver. 2.0, produced by Amersham United States Biochemical Co.). DNA sequence of the AK inserted into the pW6A is shown in SEQ ID NO: 3 and amino acids sequence coded by said DNA sequence is shown in SEQ ID NO: 4. pW6AK was introduced into host *Escherichia coli* and then cultured for 2 hours in the LB medium. Thereafter, 1 mM IPTG was added thereto, and the mixture was cultured for 2 hours to induce expression. The TE buffer were added to the precipitates of *Escherichia coli*, the precipitates were sonicated, and 15 % SDS-PAGE according to the Laemmli method was carried out. By CBB staining, a band was confirmed at about 40 Kda.

Example 13 Purification of AK

pW6AK prepared in Example 12 was introduced into host *Escherichia coli* and then cultured under conditions of using the LB medium at 37 °C. By preculture, a concentration of *Escherichia coli* in culture broth was made to have such turbidity that absorbance at a wavelength of 600 nm was about 1.0, 1 mM IPTG was added thereto to induce expression. After the mixture was cultured for 3 hours, centrifugation was carried out to recover *Escherichia coli*. 200 ml of the Tris buffer was added to recover *Escherichia coli*, followed by sonication treatment under ice cooling. After centrifugation, the expressed fused protein was recovered in the supernatant as a soluble component. When this supernatant was subjected to heat treatment at 65 °C for 10 minutes, about 70 % of the *Escherichia coli* protein was thermally denatured and precipitated, and 80 % or more of AK was recovered in the centrifugation supernatant after the heat treatment.

This supernatant was purified by a Hydroxy apatite column (manufactured by Bio-rad Lab.) equilibrated with the Tris buffer. When the supernatant was eluted by a sodium phosphate buffer, AK was recovered at a concentration of about 0.2 M sodium phosphate-eluted fraction. Then, this AK fraction was purified by gel filtration using a Superdex 200 26/60 column (trade name, manufactured by Pharmacia Biotec Co.) equilibrated with a buffer containing 6 M urea, 0.5 M sodium chloride and 20 mM Tris-hydrochloride having pH 9.4. At a fraction of a molecular weight being about 20,000, purified AK was recovered.

Example 14 Preparation of protein in which AK and Tp15 are fused

A DNA fragment of Tp15 amplified in Reference example 3 was integrated into the AK-expressing type vector pW6AK prepared in Example 12 to obtain a vector pW6AKTp15 expressing a protein in which AK and Tp15 were fused (hereinafter referred to as "AK-15" in the specification). DNA sequence of the AK-15 inserted into the vector is shown in SEQ ID NO: 20 and amino acids sequence coded by said DNA sequence is shown in SEQ ID NO: 21. In the same manner as in Example 1, the vector was introduced into *Escherichia coli*, and expression of AK-15 was induced. A sample for electrophoresis was prepared under the same conditions as in Example 1. After subjecting to 12.5 % SDS-PAGE according to the Laemmli method, one sheet of gel was subjected to CBB staining, and the other sheet was transferred to a nitrocellulose membrane by the method shown in Example 3. By using an anti-Tp15 monoclonal antibody as a primary antibody and a POD-labeled mouse IgG as a secondary antibody, these were reacted in the same method as in Example 3 and coloring was carried out by using 4-chloro-1-naphthol and hydrogen peroxide, a band was given at about 40 Kda which was the same position as that of the CBB-stained gel.

Example 15 Purification of AK-15 by heat treatment

pWAKTp15 prepared in Example 14 was introduced into host *Escherichia coli* and then cultured under conditions of using 1 liter of the LB medium at 37 °C. By preculture, a concentration of *Escherichia coli* in culture broth was made to have such turbidity that absorbance at a wavelength of 600 nm was about 1.0, 1 mM IPTG was added thereto to induce expression. The cells were recovered by centrifugation. 200 ml of a 50 mM glycine-sodium hydroxide buffer having pH 10.0 was added to the cells, and the cells were sonicated to recover AK-15 in the centrifugation supernatant. Then, by using a hot plate, heat treatment at 60 °C for 10 minutes was carried out to recover AK-15 in the centrifugation supernatant. The supernatant subjected to heat treatment was dialyzed to a 4 M urea-50 mM sodium acetate buffer having pH 6.0 and then purified by a SFF cation exchange column (trade name, manufactured by Pharmacia Biotec Co.) equilibrated with said buffer. When the supernatant was eluted by a column equilibrated buffer containing sodium chloride, AK-15 was recovered at a concentration of about 0.2 M to 0.4 M sodium chloride-eluted fraction. The recovered AK-15 fraction was purified by gel filtration using a Superdex 200 26/60 column (trade name, manufactured by Pharmacia Biotec Co.) equilibrated with a buffer containing 6 M urea, 0.5 M sodium chloride and 20 mM Tris-hydrochloride having pH 9.4. At a fraction of a molecular weight being about 40,000, purified AK-15 was recovered.

When the western blotting method was carried out in the same manner as in Example 1 by using an anti-Tp rabbit antibody, it was shown that even though purification by heat treatment was carried out, AK-15 retained reactivity. The results are shown in Fig. 12.

According to the present invention, a fused DNA sequence having more excellent operatability and productivity than those of a conventional DNA sequence coding a fused protein, a fused protein expressed from said fused DNA sequence, and a method for expressing the fused protein by using said DNA sequence.

RAW SEQUENCE LISTING
PATENT APPLICATION

SEQUENCE LISTING

(1) GENERAL INFORMATION:
(i) APPLICANT: Nobuyuki FUJII et al.
(ii) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN
(iii) NUMBER OF SEQUENCES:
(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: c/o FUJIREBIO INC., 7-1
(B) STREET: Nishi-shinjuku 2-chome
(C) CITY: Shinjuku-ku
(D) STATE: Tokyo
(E) COUNTRY: JAPAN
(F) ZIP: 163-07
(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: MS-DOS v.5
(D) SOFTWARE: Word Perfect 5.1
(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: JP 352225/1995
(B) FILING DATE: 28-DEC-1995
(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME:
(B) REGISTRATION NUMBER:
(C) REFERENCE/DOCKET NUMBER:
(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE:
(B) TELEFAX:

(2) INFORMATION FOR SEQ ID NO: 1:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 291 nucleic acids
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double strand
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthesized
(x) PUBLICATION INFORMATION:
(A) AUTHORS: Nobuyuki FUJII et al.
(B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN
(K) RELEVANT RESIDUES IN SEQ ID NO:1: FROM 1 to 291

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
ATGGCGTGGA AGGTTTCTGT CGACCAAGAC ACCTGTATAG GAGATGCCAT CTGTGCAAGC
                                30                                  60
CTCTGTCCAG ACGTCTTTGA GATGAACGAT GAAGGAAAGG CCCAACCAAA GGTAGAGGTT
                                90                                 120
ATTGAGGACG AAGAGCTCTA CAACTGTGCT AAGGAAGCTA TGGAGGCCTG TCCAGTTAGT
```

GCTATTACTA TTGAGGAGGC TGGTGGTTCT TCTCTGGTTC CGCGTGGATC GGAATTCGTC

GACCTCGAGG GATCCGGGCC CTCTAGATGC GGCCGCATGC ATGGTACCTA A

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 97 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE:
(A) DESCRIPTION: protein
(vi) ORIGINAL SOURCE:
(A) ORGANISM: recombinant
(x) PUBLICATION INFORMATION:
(A) AUTHORS: Nobuyuki FUJII et al.
(B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN
(K) RELEVANT RESIDUES IN SEQ ID NO:2: FROM 1 to 97

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Ala Trp Lys Val Ser Val Asp Gln Asp Thr Cys Ile Gly Asp Ala Ile Cys Ala Ser
 1                                                                           20
Leu Cys Pro Asp Val Phe Glu Met Asn Asp Glu Gly Lys Ala Gln Pro Lys Val Glu Val
 21                                                                          40
Ile Glu Asp Glu Glu Leu Tyr Asn Cys Ala Lys Glu Ala Met Glu Ala Cys Pro Val Ser
 41                                                                          60
Ala Ile Thr Ile Glu Glu Ala Gly Gly Ser Ser Leu Val Pro Arg Gly Ser Glu Phe Val
 61                                                                          80
Asp Leu Glu Gly Ser Gly Pro Ser Arg Cys Gly Arg Met His Gly Thr ***
 81
```

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 672 nucleic acids
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double strand
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid
(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthesized
(x) PUBLICATION INFORMATION:
(A) AUTHORS: Nobuyuki FUJII et al.
(B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN
(K) RELEVANT RESIDUES IN SEQ ID NO:3: FROM 1 to 672

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
ATGAAGATTG GTATTGTAAC TGGTATCCCT GGTGTAGGGA AAAGTACTGT CTTGGCTAAA
                                  30                                  60
GTTAAAGAGA TATTGGATAA TCAAGGTATA AATAACAAGA TCATAAATTA TGGAGATTTT
                                  90                                 120
ATGTTAGCAA CAGCATTAAA ATTAGGCTAT GCTAAAGATA GAGACGAAAT GAGAAAATTA
                                 150                                 180
TCTGTAGAAA AGCAGAAGAA ATTGCAGATT GATGCGGCTA AAGGTATAGC TGAAGAGGCA
```

```
                                    210                                 240
AGAGCAGGTG GAGAAGGATA TCTGTTCATA GATACGCACG CTGTGATACG TACACCCTCT
                                    270                                 300
GGATATTTAC CTGGTTTACC GTCAGATATA ATTACAGAAA TAAATCCGTC TGTTATCTTT
                                    330                                 360
TTACTGGAAG CTGATCCTAA GATAATATTA TCAAGGCAAA AGAGAGATAC AACAAGGAAT
                                    390                                 420
AGAAATGATT ATAGTGACGA ATCAGTTATA TTAGAAACCA TAAACTTCGC TAGATATGCA
                                    450                                 480
GCTACTGCTT CTGCAGTATT AGCCGGTTCT ACTGTTAAGG TAATTGTAAA CGTGGAAGGA
                                    510                                 540
GATCCTAGTA TAGCAGCTAA TGAGATAATA AGGTCTATGA AGGGTGGTTC TTCTCTGGTT
                                    570                                 600
CCGCGTGGAC TGGAATTCGT CGACCTCGAG GGATCCGGGC CCTCTAGATG CGGCCGCATG
                                    630                                 660
CATGGTACCT AA
```

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 224 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE:
- (A) DESCRIPTION: protein

(vi) ORIGINAL SOURCE:
- (A) ORGANISM: recombinant

(x) PUBLICATION INFORMATION:
- (A) AUTHORS: Nobuyuki FUJII et al.
- (B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN
- (K) RELEVANT RESIDUES IN SEQ ID NO:4: FROM 1 to 224

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Met Lys Ile Gly Ile Val Thr Gly Ile Pro Gly Val Gly Lys Ser Thr Val Leu Ala Lys
  1                                                                          20
Val Lys Glu Ile Leu Asp Asn Gln Gly Ile Asn Asn Lys Ile Ile Asn Tyr Gly Asp Phe
 21                                                                          40
Met Leu Ala Thr Ala Leu Lys Leu Gly Tyr Ala Lys Asp Arg Asp Glu Met Arg Lys Leu
 41                                                                          60
Ser Val Glu Lys Gln Lys Lys Leu Gln Ile Asp Ala Ala Lys Gly Ile Ala Glu Glu Ala
 61                                                                          80
Arg Ala Gly Gly Glu Gly Tyr Leu Phe Ile Asp Thr His Ala Val Ile Arg Thr Pro Ser
 81                                                                         100
Gly Tyr Leu Pro Gly Leu Pro Ser Asp Ile Ile Thr Glu Ile Asn Pro Ser Val Ile Phe
101                                                                         120
Leu Leu Glu Ala Asp Pro Lys Ile Ile Leu Ser Arg Gln Lys Arg Asp Thr Thr Arg Asn
121                                                                         140
Arg Asn Asp Tyr Ser Asp Glu Ser Val Ile Leu Glu Thr Ile Asn Phe Ala Arg Tyr Ala
141                                                                         160
Ala Thr Ala Ser Ala Val Leu Ala Gly Ser Thr Val Lys Val Ile Val Asn Val Glu Gly
161                                                                         180
Asp Pro Ser Ile Ala Ala Asn Glu Ile Ile Arg Ser Met Lys Gly Gly Ser Ser Leu Val
181                                                                         200
Pro Arg Gly Leu Glu Phe Val Asp Leu Glu Gly Ser Gly Pro Ser Arg Cys Gly Arg Met
201                                                                         220
His Gly Thr ***
221
```

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4557 nucleic acids
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double strand
(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: other nucleic acid

(vi) ORIGINAL SOURCE:
(A) ORGANISM: *E. coli*
(B) STRAIN: BL21(DE3)

(x) PUBLICATION INFORMATION:
(A) AUTHORS: Nobuyuki FUJII et al.
(B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN
(K) RELEVANT RESIDUES IN SEQ ID NO:5: FROM 1 to 4557

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
ATGGCTAGCG AATTCGTCGA CCTCGAGGGA TCCGGGCCCT CTAGATGCGG CCGCATGCAT
                                   30                              60
GGTACCTAAC TAACTAAGCT TGAGTATTCT ATAGTGTCAC CTAAATCCCA GCTTGATCCG
                                   90                             120
GCTGCTAACA AAGCCCGAAA GGAAGCTGAG TTGGCTGCTG CCACCGCTGA GCAATAACTA
                                  150                             180
GCATAACCCC TTGGGGCCTC TAAACGGGTC TTGAGGGGTT TTTTGCTGAA AGGAGGAACT
                                  210                             240
ATATCCGGAT AACCTGGCGT AATAGCGAAG AGGCCCGCAC CGAATTAATT CATCGTGACT
                                  270                             300
GACTGACGAT CTGCCTCGCG CGTTTCGGTG ATGACGGTGA AAACCTCTGA CACATGCAGC
                                  330                             360
TCCCGGAGAC GGTCACAGCT TGTCTGTAAG CGGATGCCGG GAGCAGACAA GCCCGTCAGG
                                  390                             420
GCGCGTCAGC GGGTGTTGGC GGGTGTCGGG GCGCAGCCAT GACCCAGTCA CGTAGCGATA
                                  450                             480
GCGGAGTGTA TAATTCTTGA AGACGAAAGG GCCTCGTGAT ACGCCTATTT TTATAGGTTA
                                  510                             540
ATGTCATGAT AATAATGGTT TCTTAGACGT CAGGTGGCAC TTTTCGGGGA AATGTGCGCG
                                  570                             600
GAACCCCTAT TTGTTTATTT TTCTAAATAC ATTCAAATAT GTATCCGCTC ATGAGACAAT
                                  630                             660
AACCCTGATA AATGCTTCAA TAATATTGAA AAAGGAAGAG TATGAGTATT CAACATTTCC
                                  690                             720
GTGTCGCCCT TATTCCCTTT TTTGCGGCAT TTTGCCTTCC TGTTTTTGCT CACCCAGAAA
                                  750                             780
CGCTGGTGAA AGTAAAAGAT GCTGAAGATC AGTTGGGTGC ACGAGTGGGT TACATCGAAC
                                  810                             840
TGGATCTCAA CAGCGGTAAG ATCCTTGAGA GTTTTCGCCC CGAAGAACGT TTTCCAATGA
                                  870                             900
TGAGCACTTT TAAAGTTCTG CTATGTGGCG CGGTATTATC CCGTGTTGAC GCCGGGCAAG
                                  930                             960
AGCAACTCGG TCGCCGCATA CACTATTCTC AGAATGACTT GGTTGAGTAC TCACCAGTCA
                                  990                            1020
CAGAAAAGCA TCTTACGGAT GGCATGACAG TAAGAGAATT ATGCAGTGCT GCCATAACCA
                                 1050                            1080
TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT CGGAGGACCG AAGGAGCTAA
                                 1110                            1140
CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG GAACCGGAGC
```

```
233                                1170                                1200
234  TGAATGAAGC CATACCAAAC GACGAGCGTG ACACCACGAT GCCTGCAGCA ATGGCAACAA
235                                1230                                1260
236  CGTTGCGCAA ACTATTAACT GGCGAACTAC TTACTCTAGC TTCCCGGCAA CAATTAATAG
237                                1290                                1320
238  ACTGGATGGA GGCGGATAAA GTTGCAGGAC CACTTCTGCG CTCGGCCCTT CCGGCTGGCT
239                                1350                                1380
240  GGTTTATTGC TGATAAATCT GGAGCCGGTG AGCGTGGGTC TCGCGGTATC ATTGCAGCAC
241                                1410                                1440
242  TGGGGCCAGA TGGTAAGCCC TCCCGTATCG TAGTTATCTA CACGACGGGG AGTCAGGCAA
243                                1470                                1500
244  CTATGGATGA ACGAAATAGA CAGATCGCTG AGATAGGTGC CTCACTGATT AAGCATTGGT
245                                1530                                1560
246  AACTGTCAGA CCAAGTTTAC TCATATATAC TTTAGATTGA TTTAAAACTT CATTTTTAAT
247                                1590                                1620
248  TTAAAAGGAT CTAGGTGAAG ATCCTTTTTG ATAATCTCAT GACCAAAATC CCTTAACGTG
249                                1650                                1680
250  AGTTTTCGTT CCACTGAGCG TCAGACCCCG TAGAAAAGAT CAAAGGATCT TCTTGAGATC
251                                1710                                1740
252  CTTTTTTTCT GCGCGTAATC TGCTGCTTGC AAACAAAAAA ACCACCGCTA CCAGCGGTGG
253                                1770                                1800
254  TTTGTTTGCC GGATCAAGAG CTACCAACTC TTTTTCCGAA GGTAACTGGC TTCAGCAGAG
255                                1830                                1860
256  CGCAGATACC AAATACTGTC CTTCTAGTGT AGCCGTAGTT AGGCCACCAC TTCAAGAACT
257                                1890                                1920
258  CTGTAGCACC GCCTACATAC CTCGCTCTGC TAATCCTGTT ACCAGTGGCT GCTGCCAGTG
259                                1950                                1980
260  GCGATAAGTC GTGTCTTACC GGGTTGGACT CAAGACGATA GTTACCGGAT AAGGCGCAGC
261                                2010                                2040
262  GGTCGGGCTG AACGGGGGGT TCGTGCACAC AGCCCAGCTT GGAGCGAACG ACCTACACCG
263                                2070                                2100
264  AACTGAGATA CCTACAGCGT GAGCTATGAG AAAGCGCCAC GCTTCCCGAA GGGAGAAAGG
265                                2130                                2160
266  CGGACAGGTA TCCGGTAAGC GGCAGGGTCG GAACAGGAGA GCGCACGAGG GAGCTTCCAG
267                                2190                                2220
268  GGGGAAACGC CTGGTATCTT TATAGTCCTG TCGGGTTTCG CCACCTCTGA CTTGAGCGTC
269                                2250                                2280
270  GATTTTTGTG ATGCTCGTCA GGGGGGCGGA GCCTATGGAA AAACGCCAGC AACGCGGCCT
271                                2310                                2340
272  TTTTACGGTT CCTGGCCTTT TGCTGGCCTT TTGCTCACAT GTTCTTTCCT GCGTTATCCC
273                                2370                                2400
274  CTGATTCTGT GGATAACCGT ATTACCGCCT TTGAGTGAGC TGATACCGCT CGCCGCAGCC
275                                2430                                2460
276  GAACGACCGA GCGCAGCGAG TCAGTGAGCG AGGAAGCGGA AGAGCGCCTG ATGCGGTATT
277                                2490                                2520
278  TTCTCCTTAC GCATCTGTGC GGTATTTCAC ACCGCATAAA TTCCGACACC ATCGAATGGT
279                                2550                                2580
280  GCAAAACCTT TCGCGGTATG GCATGATAGC GCCCGGAAGA GAGTCAATTC AGGGTGGTGA
281                                2610                                2640
282  ATGTGAAACC AGTAACGTTA TACGATGTCG CAGAGTATGC CGGTGTCTCT TATCAGACCG
283                                2670                                2700
284  TTTCCCGCGT GGTGAACCAG GCCAGCCACG TTTCTGCGAA AACGCGGGAA AAAGTGGAAG
285                                2730                                2760
286  CGGCGATGGC GGAGCTGAAT TACATTCCCA ACCGCGTGGC ACAACAACTG GCGGGCAAAC
287                                2790                                2820
288  AGTCGTTGCT GATTGGCGTT GCCACCTCCA GTCTGGCCCT GCACGCGCCG TCGCAAATTG
289                                2850                                2880
290  TCGCGGCGAT TAAATCTCGC GCCGATCAAC TGGGTGCCAG CGTGGTGGTG TCGATGGTAG
```

```
                                        2910                                2940
AACGAAGCGG CGTCGAAGCC TGTAAAGCGG CGGTGCACAA TCTTCTCGCG CAACGCGTCA
                                        2970                                3000
GTGGGCTGAT CATTAACTAT CCGCTGGATG ACCAGGATGC CATTGCTGTG GAAGCTGCCT
                                        3030                                3060
GCACTAATGT TCCGGCGTTA TTTCTTGATG TCTCTGACCA GACACCCATC AACAGTATTA
                                        3090                                3120
TTTTCTCCCA TGAAGACGGT ACGCGACTGG GCGTGGAGCA TCTGGTCGCA TTGGGTCACC
                                        3150                                3180
AGCAAATCGC GCTGTTAGCG GGCCCATTAA GTTCTGTCTC GGCGCGTCTG CGTCTGGCTG
                                        3210                                3240
GCTGGCATAA ATATCTCACT CGCAATCAAA TTCAGCCGAT AGCGGAACGG GAAGGCGACT
                                        3270                                3300
GGAGTGCCAT GTCCGGTTTT CAACAAACCA TGCAAATGCT GAATGAGGGC ATCGTTCCCA
                                        3330                                3360
CTGCGATGCT GGTTGCCAAC GATCAGATGG CGCTGGGCGC AATGCGCGCC ATTACCGAGT
                                        3390                                3420
CCGGGCTGCG CGTTGGTGCG GATATCTCGG TAGTGGGATA CGACGATACC GAAGACAGCT
                                        3450                                3480
CATGTTATAT CCCGCCGTTA ACCACCATCA AACAGGATTT TCGCCTGCTG GGGCAAACCA
                                        3510                                3540
GCGTGGACCG CTTGCTGCAA CTCTCTCAGG GCCAGGCGGT GAAGGGCAAT CAGCTGTTGC
                                        3570                                3600
CCGTCTCACT GGTGAAAAGA AAAACCACCC TGGCGCCCAA TACGCAAACC GCCTCTCCCC
                                        3630                                3660
GCGCGTTGGC CGATTCATTA ATGCAGCTGG CACGACAGGT TTCCCGACTG GAAAGCGGGC
                                        3690                                3720
AGTGAGCGCA ACGCAATTAA TGTGAGTTAG CTCACTCATT AGGCACCCCA GGCTTTACAC
                                        3750                                3780
TTTATGCTTC CGGCTCGTAT GTTGTGTGGA ATTGTGAGCG GATAACAATT TCACACAGGA
                                        3810                                3840
AACAGCTATG ACCATGATTA CGGATTCACT GGCCGTCGTT TTACAACGTC GTGACTGGGA
                                        3870                                3900
AAACCCTGGC GTTACCCAAC TTAATCGCCT TGCAGCACAT CCCCCTTTCG CCAGCTGGCG
                                        3930                                3960
TAATAGCGAA GAGGCCCGCA CCGATCGCCC TTCCCAACAG TTGCGCAGCC TGAATGGCGA
                                        3990                                4020
ATGGCGCTTT GCCTGGTTTC CGGCACCAGA AGCGGTGCCG GAAAGCTGGC TGGAGTGCGA
                                        4050                                4080
TCTTCCTGAG GCCGATACTG TCGTCGTCCC CTCAAACTGG CAGATGCACG GTTACGATGC
                                        4110                                4140
GCCCATCTAC ACCAACGTAA CCTATCCCAT TACGGTCAAT CCGCCGTTTG TTCCCACGGA
                                        4170                                4200
GAATCCGACG GGTTGTTACT CGCTCACATT TAATGTTGAT GAAAGCTGGC TACAGGAAGG
                                        4230                                4260
CCAGACGCGA ATTATTTTTG ATGGCGTTGG AATTACGTTA TCGACTGCAC GGTGCACCAA
                                        4290                                4320
TGCTTCTGGC GTCAGGCAGC CATCGGAAGC TGTGGTATGG CTGTGCAGGT CGTAAATCAC
                                        4350                                4380
TGCATAATTC GTGTCGCTCA AGGCGCACTC CCGTTCTGGA TAATGTTTTT TGCGCCGACA
                                        4410                                4440
TCATAACGGT TCTGGCAAAT GGGAATTGGG AAATTAATAC GACTCACTAT ATGGAATTGT
                                        4470                                4500
GAGCGGATAA CAATTCCTAG AAATAATTTT GTTTAACTTT AAGAAGGAGA TATACAT
                                        4530
```

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 672 nucleic acids
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double strand
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid, genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthesized, HTLV-I
(x) PUBLICATION INFORMATION:
(A) AUTHORS: Nobuyuki FUJII et al.
(B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN
(K) RELEVANT RESIDUES IN SEQ ID NO:6: FROM 1 to 672

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
ATGGCGTGGA AGGTTTCTGT CGACCAAGAC ACCTGTATAG GAGATGCCAT CTGTGCAAGC
                                 30                                  60
CTCTGTCCAG ACGTCTTTGA GATGAACGAT GAAGGAAAGG CCCAACCAAA GGTAGAGGTT
                                 90                                 120
ATTGAGGACG AAGAGCTCTA CAACTGTGCT AAGGAAGCTA TGGAGGCCTG TCCAGTTAGT
                                150                                 180
GCTATTACTA TTGAGGAGGC TGGTGGTTCT TCTCTGGTTC CGCGTGGATC GGAATTCATG
                                210                                 240
GGCCAAATCT TTTCCCGTAG CGCTAGCCCT ATTCCGCGGC CGCCCCGGGG GCTGGCCGCT
                                270                                 300
CATCACTGGC TTAACTTCCT CCAGGCGGCA TATCGCCTAG AACCCGGTCC CTCCAGTTAC
                                330                                 360
GATTTCCACC AGTTAAAAAA ATTTCTTAAA ATAGCTTTAG AAACACCGGT CTGGATCTGC
                                390                                 420
CCCATTAACT ACTCCCTCCT AGCCAGCCTA CTCCCAAAAG GATACCCCGG CCGGGTGAAT
                                450                                 480
GAAATTTTAC ACATACTCAT CCAAACCCAA GCCCAGATCC CGTCCCGCCC CGCGCCGCCG
                                510                                 540
CCGCCGTCAT CCTCCACCCA CGACCCCCCG GATTCTGACC CACAAATCCC CCCTCCCTAT
                                570                                 600
GTTGAGCCTA CAGCCCCCCA AGTCCTTTAA GGATCCGGGC CCTCTAGATG CGGCCGCATG
                                630                                 660
CATGGTACCT AA
```

(2) INFORMATION FOR SEQ ID NO: 7:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 224 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: protein
(vi) ORIGINAL SOURCE:
(A) ORGANISM: recombinant
(x) PUBLICATION INFORMATION:
(A) AUTHORS: Nobuyuki FUJII et al.
(B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN
(K) RELEVANT RESIDUES IN SEQ ID NO:7: FROM 1 to 224

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
Met Ala Trp Lys Val Ser Val Asp Gln Asp Thr Cys Ile Gly Asp Ala Ile Cys Ala Ser
 1                                                                              20
Leu Cys Pro Asp Val Phe Glu Met Asn Asp Glu Gly Lys Ala Gln Pro Lys Val Glu Val
```

21 40
Ile Glu Asp Glu Glu Leu Tyr Asn Cys Ala Lys Glu Ala Met Glu Ala Cys Pro Val Ser
41 60
Ala Ile Thr Ile Glu Glu Ala Gly Gly Ser Ser Leu Val Pro Arg Gly Ser Glu Phe Met
61 80
Gly Gln Ile Phe Ser Arg Ser Ala Ser Pro Ile Pro Arg Pro Pro Arg Gly Leu Ala Ala
81 100
His His Trp Leu Asn Phe Leu Gln Ala Ala Tyr Arg Leu Glu Pro Gly Pro Ser Ser Tyr
101 120
Asp Phe His Gln Leu Lys Lys Phe Leu Lys Ile Ala Leu Glu Thr Pro Val Trp Ile Cys
121 140
Pro Ile Asn Tyr Ser Leu Leu Ala Ser Leu Leu Pro Lys Gly Tyr Pro Gly Arg Val Asn
141 160
Glu Ile Leu His Ile Leu Ile Gln Thr Gln Ala Gln Ile Pro Ser Arg Pro Ala Pro Pro
161 180
Pro Pro Ser Ser Ser Thr His Asp Pro Pro Asp Ser Asp Pro Gln Ile Pro Pro Pro Tyr
181 200
Val Glu Pro Thr Ala Pro Gln Val Leu *** Gly Ser Gly Pro Ser Arg Cys Gly Arg Met
201 220
His Gly Thr ***
221

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 690 nucleic acids
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double strand
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid, genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthesized, HTLV-II

(x) PUBLICATION INFORMATION:
(A) AUTHORS: Nobuyuki FUJII et al.
(B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN
(K) RELEVANT RESIDUES IN SEQ ID NO:8: FROM 1 to 690

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

ATGGCGTGGA AGGTTTCTGT CGACCAAGAC ACCTGTATAG GAGATGCCAT CTGTGCAAGC
30 60
CTCTGTCCAG ACGTCTTTGA GATGAACGAT GAAGGAAAGG CCCAACCAAA GGTAGAGGTT
90 120
ATTGAGGACG AAGAGCTCTA CAACTGTGCT AAGGAAGCTA TGGAGGCCTG TCCAGTTAGT
150 180
GCTATTACTA TTGAGGAGGC TGGTGGTTCT TCTCTGGTTC CGCGTGGATC GGAATTCATG
210 240
GGACAAATCC ACGGGCTTTC CCCAACTCCA ATACCCAAAG CCCCCAGGGG GCTATCAACC
270 300
CACCACTGGC TTAACTTTCT CCAGGCTGCT TACCGCTTGC AGCCTAGGCC CTCCGATTTC
330 360
GACTTCCAGC AGCTACGACG CTTTCTAAAA CTAGCCCTTA AAACGCCCAT TTGGCTAAAT
390 420
CCTATTGACT ACTCGCTTTT AGCTAGCCTT ATCCCCAAGG GATATCCAGG AAGGGTGGTA
450 480
GAGATTATAA ATATCCTTGT CAAAAATCAA GTCTCCCCTA GCGCCCCCGC CGCCCCAGTT
510 540
CCGACACCTA TCTGCCCTAC TACTACTCCT CCGCCACCTC CCCCCCCTTC CCCGGAGGCC

570 600
CATGTTCCCC CCCCTTACGT GGAACCCACC ACCACGCAAT GCTTCTAAGG ATCCGGGCCC
630 660
TCTAGATGCG GCCGCATGCA TGGTACCTAA
690

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 230 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: protein
(vi) ORIGINAL SOURCE:
(A) ORGANISM: recombinant
(x) PUBLICATION INFORMATION:
(A) AUTHORS: Nobuyuki FUJII et al.
(B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN
(K) RELEVANT RESIDUES IN SEQ ID NO:9: FROM 1 to 230

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met Ala Trp Lys Val Ser Val Asp Gln Asp Thr Cys Ile Gly Asp Ala Ile Cys Ala Ser
1 20
Leu Cys Pro Asp Val Phe Glu Met Asn Asp Glu Gly Lys Ala Gln Pro Lys Val Glu Val
21 40
Ile Glu Asp Glu Glu Leu Tyr Asn Cys Ala Lys Glu Ala Met Glu Ala Cys Pro Val Ser
41 60
Ala Ile Thr Ile Glu Glu Ala Gly Gly Ser Ser Leu Val Pro Arg Gly Ser Glu Phe Met
61 80
Gly Gln Ile His Gly Leu Ser Pro Thr Pro Ile Pro Lys Ala Pro Arg Gly Leu Ser Thr
81 100
His His Trp Leu Asn Phe Leu Gln Ala Ala Tyr Arg Leu Gln Pro Arg Pro Ser Asp Phe
101 120
Asp Phe Gln Gln Leu Arg Arg Phe Leu Lys Leu Ala Leu Lys Thr Pro Ile Trp Leu Asn
121 140
Pro Ile Asp Tyr Ser Leu Leu Ala Ser Leu Ile Pro Lys Gly Tyr Pro Gly Arg Val Val
141 160
Glu Ile Ile Asn Ile Leu Val Lys Asn Gln Val Ser Pro Ser Ala Pro Ala Ala Pro Val
161 180
Pro Thr Pro Ile Cys Pro Thr Thr Thr Pro Pro Pro Pro Pro Pro Pro Ser Pro Glu Ala
181 200
His Val Pro Pro Pro Tyr Val Glu Pro Thr Thr Thr Gln Cys Phe *** Gly Ser Gly Pro
201 220
Ser Arg Cys Gly Arg Met His Gly Thr ***
221

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 810 nucleic acids
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double strand
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid, genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthesized, HTLV-I
(x) PUBLICATION INFORMATION:
(A) AUTHORS: Nobuyuki FUJII et al.

(B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN
(K) RELEVANT RESIDUES IN SEQ ID NO:10: FROM 1 to 810

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
ATGGCGTGGA AGGTTTCTGT CGACCAAGAC ACCTGTATAG GAGATGCCAT CTGTGCAAGC
                                 30                                  60
CTCTGTCCAG ACGTCTTTGA GATGAACGAT GAAGGAAAGG CCCAACCAAA GGTAGAGGTT
                                 90                                 120
ATTGAGGACG AAGAGCTCTA CAACTGTGCT AAGGAAGCTA TGGAGGCCTG TCCAGTTAGT
                                150                                 180
GCTATTACTA TTGAGGAGGC TGGTGGTTCT TCTCTGGTTC CGCGTGGATC GGAATTCGCA
                                210                                 240
GTACCGGTGG CGGTCTGGCT TGTCTCCGCC CTGGCCATGG GAGCCGGAGT GGCTGGCAGG
                                270                                 300
ATTACCGGCT CCATGTCCCT CGCCTCAGGA AAGAGCCTCC TACATGAGGT GGACAAAGAT
                                330                                 360
ATTTCCCAAT TAACTCAAGC AATAGTCAAA AACCACAAAA ATCTGCTCAA AATTGCACAG
                                390                                 420
TATGCTGCCC AGAACAGACG AGGCCTTGAT CTCCTGTTCT GGGAGCAAGG AGGATTATGC
                                450                                 480
AAAGCATTAC AAGAACAGTG CTGTTTTCTA AATATTACTA ATTCCCATGT CTCAATACTA
                                510                                 540
CAAGAGAGAC CCCCCCTTGA AAATCGAGTC CTGACTGGCT GGGGCCTTAA CTGGGACCTT
                                570                                 600
GGCCTCTCAC AGTGGGCTCG AGAAGCCTTA CAAACTGGAA TCACCCTTGT CGCGCTACTC
                                630                                 660
CTTCTTGTTA TCCTTGCAGG ACCATGCATC CTCCGTCAGC TACGACACCT CCCCTCGCGC
                                690                                 720
GTCAGATACC CCCATTACTC TCTTATAAAC CCTGAGTCAT CCCTGTAAGG ATCCGGGCCC
                                750                                 780
TCTAGATGCG GCCGCATGCA TGGTACCTAA
                                810
```

(2) INFORMATION FOR SEQ ID NO: 11:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 270 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: protein
(vi) ORIGINAL SOURCE:
(A) ORGANISM: recombinant
(x) PUBLICATION INFORMATION:
(A) AUTHORS: Nobuyuki FUJII et al.
(B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN
(K) RELEVANT RESIDUES IN SEQ ID NO:11: FROM 1 to 270

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
Met Ala Trp Lys Val Ser Val Asp Gln Asp Thr Cys Ile Gly Asp Ala Ile Cys Ala Ser
 1                                                                            20
Leu Cys Pro Asp Val Phe Glu Met Asn Asp Glu Gly Lys Ala Gln Pro Lys Val Glu Val
 21                                                                           40
Ile Glu Asp Glu Glu Leu Tyr Asn Cys Ala Lys Glu Ala Met Glu Ala Cys Pro Val Ser
 41                                                                           60
Ala Ile Thr Ile Glu Glu Ala Gly Gly Ser Ser Leu Val Pro Arg Gly Ser Glu Phe Ala
```

```
                61                                                              80
Val Pro Val Ala Val Trp Leu Val Ser Ala Leu Ala Met Gly Ala Gly Val Ala Gly Arg
                81                                                             100
Ile Thr Gly Ser Met Ser Leu Ala Ser Gly Lys Ser Leu Leu His Glu Val Asp Lys Asp
               101                                                             120
Ile Ser Gln Leu Thr Gln Ala Ile Val Lys Asn His Lys Asn Leu Leu Lys Ile Ala Gln
               121                                                             140
Tyr Ala Ala Gln Asn Arg Arg Gly Leu Asp Leu Leu Phe Trp Glu Gln Gly Gly Leu Cys
               141                                                             160
Lys Ala Leu Gln Glu Gln Cys Cys Phe Leu Asn Ile Thr Asn Ser His Val Ser Ile Leu
               161                                                             180
Gln Glu Arg Pro Pro Leu Glu Asn Arg Val Leu Thr Gly Trp Gly Leu Asn Trp Asp Leu
               181                                                             200
Gly Leu Ser Gln Trp Ala Arg Glu Ala Leu Gln Thr Gly Ile Thr Leu Val Ala Leu Leu
               201                                                             220
Leu Leu Val Ile Leu Ala Gly Pro Cys Ile Leu Arg Gln Leu Arg His Leu Pro Ser Arg
               221                                                             240
Val Arg Tyr Pro His Tyr Ser Leu Ile Asn Pro Glu Ser Ser Leu *** Gly Ser Gly Pro
               241                                                             260
Ser Arg Cys Gly Arg Met His Gly Thr ***
               261

```

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 816 nucleic acids
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: double strand
  (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid, genomic DNA

(vi) ORIGINAL SOURCE:
  (A) ORGANISM: synthesized, HTLV-II

(x) PUBLICATION INFORMATION:
  (A) AUTHORS: Nobuyuki FUJII et al.
  (B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN
  (K) RELEVANT RESIDUES IN SEQ ID NO:12: FROM 1 to 816

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
ATGGCGTGGA AGGTTTCTGT CGACCAAGAC ACCTGTATAG GAGATGCCAT CTGTGCAAGC
                                  30                                  60
CTCTGTCCAG ACGTCTTTGA GATGAACGAT GAAGGAAAGG CCCAACCAAA GGTAGAGGTT
                                  90                                 120
ATTGAGGACG AAGAGCTCTA CAACTGTGCT AAGGAAGCTA TGGAGGCCTG TCCAGTTAGT
                                 150                                 180
GCTATTACTA TTGAGGAGGC TGGTGGTTCT TCTCTGGTTC CGCGTGGATC GGAATTCGCC
                                 210                                 240
GTTCCAATAG CAGTGTGGCT TGTCTCCGCC CTAGCGGCCG GAACAGGTAT CGCTGGTGGA
                                 270                                 300
GTAACAGGCT CCCTATCTCT GGCTTCCAGT AAAAGCCTTC TCCTCGAGGT TGACAAAGAC
                                 330                                 360
ATCTCCCACC TTACCCAGGC CATAGTCAAA AATCATCAAA ACATCCTCCG GGTTGCACAG
                                 390                                 420
TATGCAGCCC AAAATAGACG AGGATTAGAC CTCCTATTCT GGGAACAAGG GGGTTTGTGC
                                 450                                 480
AAGGCCATAC AGGAGCAATG TTGCTTCCTC AACATCAGTA ACACTCATGT ATCCGTCCTC
                                 510                                 540
CAGGAACGGC CCCCTCTTGA AAAACGTGTC ATCACCGGCT GGGGACTAAA CTGGGATCTT
```

570 600
GGACTGTCCC AATGGGCACG AGAAGCCCTC CAGACAGGCA TAACCATTCT CGCTCTACTC
630 660
CTCCTCGTCA TATTGTTTGG CCCCTGTATC CTCCGCCAAA TCCAGGCCCT TCCACAGCGG
690 720
TTACAAAACC GACATAACCA GTATTCCCTT ATCAACCCAG AAACCATGCT ATAAGGATCC
750 780
GGGCCCTCTA GATGCGGCCG CATGCATGGT ACCTAA
810

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 272 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
(A) ORGANISM: recombinant

(x) PUBLICATION INFORMATION:
(A) AUTHORS: Nobuyuki FUJII et al.
(B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN
(K) RELEVANT RESIDUES IN SEQ ID NO:13: FROM 1 to 272

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Met Ala Trp Lys Val Ser Val Asp Gln Asp Thr Cys Ile Gly Asp Ala Ile Cys Ala Ser
1 20
Leu Cys Pro Asp Val Phe Glu Met Asn Asp Glu Gly Lys Ala Gln Pro Lys Val Glu Val
21 40
Ile Glu Asp Glu Glu Leu Tyr Asn Cys Ala Lys Glu Ala Met Glu Ala Cys Pro Val Ser
41 60
Ala Ile Thr Ile Glu Glu Ala Gly Gly Ser Ser Leu Val Pro Arg Gly Ser Glu Phe Ala
61 80
Val Pro Ile Ala Val Trp Leu Val Ser Ala Leu Ala Ala Gly Thr Gly Ile Ala Gly Gly
81 100
Val Thr Gly Ser Leu Ser Leu Ala Ser Ser Lys Ser Leu Leu Leu Glu Val Asp Lys Asp
101 120
Ile Ser His Leu Thr Gln Ala Ile Val Lys Asn His Gln Asn Ile Leu Arg Val Ala Gln
121 140
Tyr Ala Ala Gln Asn Arg Arg Gly Leu Asp Leu Leu Phe Trp Glu Gln Gly Gly Leu Cys
141 160
Lys Ala Ile Gln Glu Gln Cys Cys Phe Leu Asn Ile Ser Asn Thr His Val Ser Val Leu
161 180
Gln Glu Arg Pro Pro Leu Glu Lys Arg Val Ile Thr Gly Trp Gly Leu Asn Trp Asp Leu
181 200
Gly Leu Ser Gln Trp Ala Arg Glu Ala Leu Gln Thr Gly Ile Thr Ile Leu Ala Leu Leu
201 220
Leu Leu Val Ile Leu Phe Gly Pro Cys Ile Leu Arg Gln Ile Gln Ala Leu Pro Gln Arg
221 240
Leu Gln Asn Arg His Asn Gln Tyr Ser Leu Ile Asn Pro Glu Thr Met Leu *** Gly Ser
241 260
Gly Pro Ser Arg Cys Gly Arg Met His Gly Thr ***
261

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1119 nucleic acids

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double strand
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid, genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: plasmid, Tp
(B) STRAIN: Nichols
(x) PUBLICATION INFORMATION:
(A) AUTHORS: Nobuyuki FUJII et al.
(B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN
(K) RELEVANT RESIDUES IN SEQ ID NO:14: FROM 1 to 1119

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```
ATGTCCCCTA TACTAGGTTA TTGGAAAATT AAGGGCCTTG TGCAACCCAC TCGACTTCTT
                                   30                                  60
TTGGAATATC TTGAAGAAAA ATATGAAGAG CATTTGTATG AGCGCGATGA AGGTGATAAA
                                   90                                 120
TGGCGAAACA AAAAGTTTGA ATTGGGTTTG GAGTTTCCCA ATCTTCCTTA TTATATTGAT
                                  150                                 180
GGTGATGTTA AATTAACACA GTCTATGGCC ATCATACGTT ATATAGCTGA CAAGCACAAC
                                  210                                 240
ATGTTGGGTG GTTGTCCAAA AGAGCGTGCA GAGATTTCAA TGCTTGAAGG AGCGGTTTTG
                                  270                                 300
GATATTAGAT ACGGTGTTTC GAGAATTGCA TATAGTAAAG ACTTTGAAAC TCTCAAAGTT
                                  330                                 360
GATTTTCTTA GCAAGCTACC TGAAATGCTG AAAATGTTCG AAGATCGTTT ATGTCATAAA
                                  390                                 420
ACATATTTAA ATGGTGATCA TGTAACCCAT CCTGACTTCA TGTTGTATGA CGCTCTTGAT
                                  450                                 480
GTTGTTTTAT ACATGGACCC AATGTGCCTG GATGCGTTCC CAAAATTAGT TTGTTTTAAA
                                  510                                 540
AAACGTATTG AAGCTATCCC ACAAATTGAT AAGTACTTGA AATCCAGCAA GTATATAGCA
                                  570                                 600
TGGCCTTTGC AGGGCTGGCA AGCCACGTTT GGTGGTGGCG ACCATCCTCC AAAATCGGAT
                                  630                                 660
CTGGTTCCGC GTGGATCGGA ATTCTGTTCA TTTAGTTCTA TCCCGAATGG CACGTACCGG
                                  690                                 720
GCGACGTATC AGGATTTTGA TGAGAATGGT TGGAAGGACT TTCTCGAGGT TACTTTTGAT
                                  750                                 780
GGTGGCAAGA TGGTGCAGGT GGTTTACGAT TATCAGCATA AAGAAGGGCG GTTTAAGTCC
                                  810                                 840
CAGGACGCTG ACTACCATCG GGTCATGTAT GCATCCTCGG GCATAGGTCC TGAAAAGGCC
                                  870                                 900
TTCAGAGAGC TCGCCGATGC TTTGCTTGAA AAGGGTAATC CCGAGATGGT GGATGTGGTC
                                  930                                 960
ACCGGTGCAA CTGTTTCTTC CCAGAGTTTC AGGAGGTTGG GTCGTGCGCT TCTGCAGAGT
                                  990                                1020
GCGCGGCGCG GCGAGAAGGA AGCCATTATT AGCAGGTAGG AATTCGTCGA CCTCGAGGGA
                                 1050                                1080
TCCGGGCCCT CTAGATGCGG CCGCATGCAT GGTACCTAA
                                 1110
```

(2) INFORMATION FOR SEQ ID NO: 15:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 373 amino acids
(B) TYPE: amino acid

(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: protein
(vi) ORIGINAL SOURCE:
(A) ORGANISM: recombinant
(x) PUBLICATION INFORMATION:
(A) AUTHORS: Nobuyuki FUJII et al.
(B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN
(K) RELEVANT RESIDUES IN SEQ ID NO:15: FROM 1 to 373

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

```
Met Ser Pro Ile Leu Gly Tyr Trp Lys Ile Lys Gly Leu Val Gln Pro Thr Arg Leu Leu
 1                                                                           20
Leu Glu Tyr Leu Glu Glu Lys Tyr Glu Glu His Leu Tyr Glu Arg Asp Glu Gly Asp Lys
 21                                                                          40
Trp Arg Asn Lys Lys Phe Glu Leu Gly Leu Glu Phe Pro Asn Leu Pro Tyr Tyr Ile Asp
 41                                                                          60
Gly Asp Val Lys Leu Thr Gln Ser Met Ala Ile Ile Arg Tyr Ile Ala Asp Lys His Asn
 61                                                                          80
Met Leu Gly Gly Cys Pro Lys Glu Arg Ala Glu Ile Ser Met Leu Glu Gly Ala Val Leu
 81                                                                         100
Asp Ile Arg Tyr Gly Val Ser Arg Ile Ala Tyr Ser Lys Asp Phe Glu Thr Leu Lys Val
101                                                                         120
Asp Phe Leu Ser Lys Leu Pro Glu Met Leu Lys Met Phe Glu Asp Arg Leu Cys His Lys
121                                                                         140
Thr Tyr Leu Asn Gly Asp His Val Thr His Pro Asp Phe Met Leu Tyr Asp Ala Leu Asp
141                                                                         160
Val Val Leu Tyr Met Asp Pro Met Cys Leu Asp Ala Phe Pro Lys Leu Val Cys Phe Lys
161                                                                         180
Lys Arg Ile Glu Ala Ile Pro Gln Ile Asp Lys Tyr Leu Lys Ser Ser Lys Tyr Ile Ala
181                                                                         200
Trp Pro Leu Gln Gly Trp Gln Ala Thr Phe Gly Gly Gly Asp His Pro Pro Lys Ser Asp
201                                                                         220
Leu Val Pro Arg Gly Ser Glu Phe Cys Ser Phe Ser Ser Ile Pro Asn Gly Thr Tyr Arg
221                                                                         240
Ala Thr Tyr Gln Asp Phe Asp Glu Asn Gly Trp Lys Asp Phe Leu Glu Val Thr Phe Asp
241                                                                         260
Gly Gly Lys Met Val Gln Val Val Tyr Asp Tyr Gln His Lys Glu Gly Arg Phe Lys Ser
261                                                                         280
Gln Asp Ala Asp Tyr His Arg Val Met Tyr Ala Ser Ser Gly Ile Gly Pro Glu Lys Ala
281                                                                         300
Phe Arg Glu Leu Ala Asp Ala Leu Leu Glu Lys Gly Asn Pro Glu Met Val Asp Val Val
301                                                                         320
Thr Gly Ala Thr Val Ser Ser Gln Ser Phe Arg Arg Leu Gly Arg Ala Leu Leu Gln Ser
321                                                                         340
Ala Arg Arg Gly Glu Lys Glu Ala Ile Ile Ser Arg *** Glu Phe Val Asp Leu Glu Gly
341                                                                         360
Ser Gly Pro Ser Arg Cys Gly Arg Met His Gly Thr ***
361
```

(2) INFORMATION FOR SEQ ID NO: 16:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 858 nucleic acids
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double strand
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: genomic DNA

(vi) ORIGINAL SOURCE:
(A) ORGANISM: *E. coli*, Tp
(B) STRAIN: DH15α, Nichols
(x) PUBLICATION INFORMATION:
(A) AUTHORS: Nobuyuki FUJII et al.
(B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN
(K) RELEVANT RESIDUES IN SEQ ID NO:16: FROM 1 to 858

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```
ATGTTACACC AACAACGAAA CCAACACGCC AGGCTTATTC CTGTGGAGTT ATATATGAGC
                                   30                              60
GATAAAATTA TTCACCTGAC TGACGACAGT TTTGACACGG ATGTACTCAA AGCGGACGGG
                                   90                             120
GCGATCCTCG TCGATTTCTG GGCAGAGTGG TGCGGTCCGT GCAAAATGAT CGCCCCGATT
                                  150                             180
CTGGATGAAA TCGCTGACGA ATATCAGGGC AAACTGACCG TTGCAAAACT GAACATCGAT
                                  210                             240
CAAAACCCTG GCACTGCGCC GAAATATGGC ATCCGTGGTA TCCCGACTCT GCTGCTGTTC
                                  270                             300
AAAAACGGTG AAGTGGCGGC AACCAAAGTG GGTGCACTGT CTAAAGGTCA GTTGAAAGAG
                                  330                             360
TTCCTCGACG CTAACCTGGC GGAGCTCGGT GGTTCTTCTC TGGTTCCGCG TGGATCGGAA
                                  390                             420
TTCTGTTCAT TTAGTTCTAT CCCGAATGGC ACGTACCGGG CGACGTATCA GGATTTTGAT
                                  450                             480
GAGAATGGTT GGAAGGACTT TCTCGAGGTT ACTTTTGATG GTGGCAAGAT GGTGCAGGTG
                                  510                             540
GTTTACGATT ATCAGCATAA AGAAGGGCGG TTTAAGTCCC AGGACGCTGA CTACCATCGG
                                  570                             600
GTCATGTATG CATCCTCGGG CATAGGTCCT GAAAAGGCCT TCAGAGAGCT CGCCGATGCT
                                  630                             660
TTGCTTGAAA AGGGTAATCC CGAGATGGTG GATGTGGTCA CCGGTGCAAC TGTTTCTTCC
                                  690                             720
CAGAGTTTCA GGAGGTTGGG TCGTGCGCTT CTGCAGAGTG CGCGGCGCGG CGAGAAGGAA
                                  750                             780
GCCATTATTA GCAGGTAGGA ATTCGTCGAC CTCGAGGGAT CCGGGCCCTC TAGATGCGGC
                                  810                             840
CGCATGCATG GTACCTAA
```

(2) INFORMATION FOR SEQ ID NO: 17:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 286 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: protein
(vi) ORIGINAL SOURCE:
(A) ORGANISM: recombinant
(x) PUBLICATION INFORMATION:
(A) AUTHORS: Nobuyuki FUJII et al.
(B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN
(K) RELEVANT RESIDUES IN SEQ ID NO:17: FROM 1 to 286

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

```
Met Leu His Gln Gln Arg Asn Gln His Ala Arg Leu Ile Pro Val Glu Leu Tyr Met Ser
 1                                                                           20
Asp Lys Ile Ile His Leu Thr Asp Asp Ser Phe Asp Thr Asp Val Leu Lys Ala Asp Gly
 21                                                                          40
Ala Ile Leu Val Asp Phe Trp Ala Glu Trp Cys Gly Pro Cys Lys Met Ile Ala Pro Ile
 41                                                                          60
Leu Asp Glu Ile Ala Asp Glu Tyr Gln Gly Lys Leu Thr Val Ala Lys Leu Asn Ile Asp
 61                                                                          80
Gln Asn Pro Gly Thr Ala Pro Lys Tyr Gly Ile Arg Gly Ile Pro Thr Leu Leu Leu Phe
 81                                                                         100
Lys Asn Gly Glu Val Ala Ala Thr Lys Val Gly Ala Leu Ser Lys Gly Gln Leu Lys Glu
101                                                                         120
Phe Leu Asp Ala Asn Leu Ala Glu Leu Gly Gly Ser Ser Leu Val Pro Arg Gly Ser Glu
121                                                                         140
Phe Cys Ser Phe Ser Ser Ile Pro Asn Gly Thr Tyr Arg Ala Thr Tyr Gln Asp Phe Asp
141                                                                         160
Glu Asn Gly Trp Lys Asp Phe Leu Glu Val Thr Phe Asp Gly Gly Lys Met Val Gln Val
161                                                                         180
Val Tyr Asp Tyr Gln His Lys Glu Gly Arg Phe Lys Ser Gln Asp Ala Asp Tyr His Arg
181                                                                         200
Val Met Tyr Ala Ser Ser Gly Ile Gly Pro Glu Lys Ala Phe Arg Glu Leu Ala Asp Ala
201                                                                         220
Leu Leu Glu Lys Gly Asn Pro Glu Met Val Asp Val Val Thr Gly Ala Thr Val Ser Ser
221                                                                         240
Gln Ser Phe Arg Arg Leu Gly Arg Ala Leu Leu Gln Ser Ala Arg Arg Gly Glu Lys Glu
241                                                                         260
Ala Ile Ile Ser Arg *** Glu Phe Val Asp Leu Glu Gly Ser Gly Pro Ser Arg Cys Gly
261                                                                         280
Arg Met His Gly Thr ***
281
```

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 672 nucleic acids
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double strand
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid, genomic DNA

(vi) ORIGINAL SOURCE:
- (A) ORGANISM: synthesized, Tp
- (B) STRAIN: Nichols

(x) PUBLICATION INFORMATION:
- (A) AUTHORS: Nobuyuki FUJII et al.
- (B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN
- (K) RELEVANT RESIDUES IN SEQ ID NO:18: FROM 1 to 672

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

```
ATGGCGTGGA AGGTTTCTGT CGACCAAGAC ACCTGTATAG GAGATGCCAT CTGTGCAAGC
                                  30                                  60
CTCTGTCCAG ACGTCTTTGA GATGAACGAT GAAGGAAAGG CCCAACCAAA GGTAGAGGTT
                                  90                                 120
ATTGAGGACG AAGAGCTCTA CAACTGTGCT AAGGAAGCTA TGGAGGCCTG TCCAGTTAGT
                                 150                                 180
GCTATTACTA TTGAGGAGGC TGGTGGTTCT TCTCTGGTTC CGCGTGGATC GGAATTCTGT
                                 210                                 240
TCATTTAGTT CTATCCCGAA TGGCACGTAC CGGGCGACGT ATCAGGATTT TGATGAGAAT
```

```
                                    270                                 300
GGTTGGAAGG ACTTTCTCGA GGTTACTTTT GATGGTGGCA AGATGGTGCA GGTGGTTTAC
                                    330                                 360
GATTATCAGC ATAAAGAAGG GCGGTTTAAG TCCCAGGACG CTGACTACCA TCGGGTCATG
                                    390                                 420
TATGCATCCT CGGGCATAGG TCCTGAAAAG GCCTTCAGAG AGCTCGCCGA TGCTTTGCTT
                                    450                                 480
GAAAAGGGTA ATCCCGAGAT GGTGGATGTG GTCACCGGTG CAACTGTTTC TTCCCAGAGT
                                    510                                 540
TTCAGGAGGT TGGGTCGTGC GCTTCTGCAG AGTGCGCGGC GCGGCGAGAA GGAAGCCATT
                                    570                                 600
ATTAGCAGGT AGGAATTCGT CGACCTCGAG GGATCCGGGC CCTCTAGATG CGGCCGCATG
                                    630                                 660
CATGGTACCT AA
```

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 224 amino acids
  (B) TYPE: amino acid
  (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:
  (A) ORGANISM: recombinant

(x) PUBLICATION INFORMATION:
  (A) AUTHORS: Nobuyuki FUJII et al.
  (B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN
  (K) RELEVANT RESIDUES IN SEQ ID NO:19: FROM 1 to 224

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

```
Met Ala Trp Lys Val Ser Val Asp Gln Asp Thr Cys Ile Gly Asp Ala Ile Cys Ala Ser
 1                                                                           20
Leu Cys Pro Asp Val Phe Glu Met Asn Asp Glu Gly Lys Ala Gln Pro Lys Val Glu Val
 21                                                                          40
Ile Glu Asp Glu Glu Leu Tyr Asn Cys Ala Lys Glu Ala Met Glu Ala Cys Pro Val Ser
 41                                                                          60
Ala Ile Thr Ile Glu Glu Ala Gly Gly Ser Ser Leu Val Pro Arg Gly Ser Glu Phe Cys
 61                                                                          80
Ser Phe Ser Ser Ile Pro Asn Gly Thr Tyr Arg Ala Thr Tyr Gln Asp Phe Asp Glu Asn
 81                                                                         100
Gly Trp Lys Asp Phe Leu Glu Val Thr Phe Asp Gly Gly Lys Met Val Gln Val Val Tyr
101                                                                         120
Asp Tyr Gln His Lys Glu Gly Arg Phe Lys Ser Gln Asp Ala Asp Tyr His Arg Val Met
121                                                                         140
Tyr Ala Ser Ser Gly Ile Gly Pro Glu Lys Ala Phe Arg Glu Leu Ala Asp Ala Leu Leu
 141                                                                        160
Glu Lys Gly Asn Pro Glu Met Val Asp Val Val Thr Gly Ala Thr Val Ser Ser Gln Ser
161                                                                         180
Phe Arg Arg Leu Gly Arg Ala Leu Leu Gln Ser Ala Arg Arg Gly Glu Lys Glu Ala Ile
181                                                                         200
Ile Ser Arg *** Glu Phe Val Asp Leu Glu Gly Ser Gly Pro Ser Arg Cys Gly Arg Met
201                                                                         220
His Gly Thr ***
221
```

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1035 nucleic acids
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double strand
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: other nucleic acid, genomic DNA
(vi) ORIGINAL SOURCE:
(A) ORGANISM: synthesized, Tp
(B) STRAIN: Nichols
(x) PUBLICATION INFORMATION:
(A) AUTHORS: Nobuyuki FUJII et al.
(B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN
(K) RELEVANT RESIDUES IN SEQ ID NO:20: FROM 1 to 1035

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

```
ATGAAGATTG GTATTGTAAC TGGTATCCCT GGTGTAGGGA AAAGTACTGT CTTGGCTAAA
                                   30                                  60
GTTAAAGAGA TATTGGATAA TCAAGGTATA AATAACAAGA TCATAAATTA TGGAGATTTT
                                   90                                 120
ATGTTAGCAA CAGCATTAAA ATTAGGCTAT GCTAAAGATA GAGACGAAAT GAGAAAATTA
                                  150                                 180
TCTGTAGAAA AGCAGAAGAA ATTGCAGATT GATGCGGCTA AAGGTATAGC TGAAGAGGCA
                                  210                                 240
AGAGCAGGTG GAGAAGGATA TCTGTTCATA GATACGCACG CTGTGATACG TACACCCTCT
                                  270                                 300
GGATATTTAC CTGGTTTACC GTCAGATATA ATTACAGAAA TAAATCCGTC TGTTATCTTT
                                  330                                 360
TTACTGGAAG CTGATCCTAA GATAATATTA TCAAGGCAAA AGAGAGATAC AACAAGGAAT
                                  390                                 420
AGAAATGATT ATAGTGACGA ATCAGTTATA TTAGAAACCA TAAACTTCGC TAGATATGCA
                                  450                                 480
GCTACTGCTT CTGCAGTATT AGCCGGTTCT ACTGTTAAGG TAATTGTAAA CGTGGAAGGA
                                  510                                 540
GATCCTAGTA TAGCAGCTAA TGAGATAATA AGGTCTATGA AGGGTGGTTC TTCTCTGGTT
                                  570                                 600
CCGCGTGGAT CGGAATTCTG TTCATTTAGT TCTATCCCGA ATGGCACGTA CCGGGCGACG
                                  630                                 660
TATCAGGATT TTGATGAGAA TGGTTGGAAG GACTTTCTCG AGGTTACTTT TGATGGTGGC
                                  690                                 720
AAGATGGTGC AGGTGGTTTA CGATTATCAG CATAAAGAAG GGCGGTTTAA GTCCCAGGAC
                                  750                                 780
GCTGACTACC ATCGGGTCAT GTATGCATCC TCGGGCATAG GTCCTGAAAA GGCCTTCAGA
                                  810                                 840
GAGCTCGCCG ATGCTTTGCT TGAAAAGGGT AATCCCGAGA TGGTGGATGT GGTCACCGGT
                                  870                                 900
GCAACTGTTT CTTCCCAGAG TTTCAGGAGG TTGGGTCGTG CGCTTCTGCA GAGTGCGCGG
                                  930                                 960
CGCGGCGAGA AGGAAGCCAT TATTAGCAGG TAGGGATCCG GGCCCTCTAG ATGCGGCCGC
                                  990                                1020
ATGCATGGTA CCTAA
```

(2) INFORMATION FOR SEQ ID NO: 21:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 345 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(vi) ORIGINAL SOURCE:
(A) ORGANISM: recombinant
(x) PUBLICATION INFORMATION:
(A) AUTHORS: Nobuyuki FUJII et al.
(B) TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED FROM SAID FUSED DNA SEQUENCE AND METHOD FOR EXPRESSING SAID FUSED PROTEIN
(K) RELEVANT RESIDUES IN SEQ ID NO:21: FROM 1 to 345

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

```
Met Lys Ile Gly Ile Val Thr Gly Ile Pro Gly Val Gly Lys Ser Thr Val Leu Ala Lys
 1                                                                           20
Val Lys Glu Ile Leu Asp Asn Gln Gly Ile Asn Asn Lys Ile Ile Asn Tyr Gly Asp Phe
 21                                                                          40
Met Leu Ala Thr Ala Leu Lys Leu Gly Tyr Ala Lys Asp Arg Asp Glu Met Arg Lys Leu
 41                                                                          60
Ser Val Glu Lys Gln Lys Lys Leu Gln Ile Asp Ala Ala Lys Gly Ile Ala Glu Glu Ala
 61                                                                          80
Arg Ala Gly Gly Glu Gly Tyr Leu Phe Ile Asp Thr His Ala Val Ile Arg Thr Pro Ser
 81                                                                         100
Gly Tyr Leu Pro Gly Leu Pro Ser Asp Ile Ile Thr Glu Ile Asn Pro Ser Val Ile Phe
 101                                                                        120
Leu Leu Glu Ala Asp Pro Lys Ile Ile Leu Ser Arg Gln Lys Arg Asp Thr Thr Arg Asn
121                                                                         140
Arg Asn Asp Tyr Ser Asp Glu Ser Val Ile Leu Glu Thr Ile Asn Phe Ala Arg Tyr Ala
141                                                                         160
Ala Thr Ala Ser Ala Val Leu Ala Gly Ser Thr Val Lys Val Ile Val Asn Val Glu Gly
161                                                                         180
Asp Pro Ser Ile Ala Ala Asn Glu Ile Ile Arg Ser Met Lys Gly Gly Ser Ser Leu Val
181                                                                         200
Pro Arg Gly Ser Glu Phe Cys Ser Phe Ser Ser Ile Pro Asn Gly Thr Tyr Arg Ala Thr
201                                                                         220
Tyr Gln Asp Phe Asp Glu Asn Gly Trp Lys Asp Phe Leu Glu Val Thr Phe Asp Gly Gly
221                                                                         240
Lys Met Val Gln Val Val Tyr Asp Tyr Gln His Lys Glu Gly Arg Phe Lys Ser Gln Asp
241                                                                         260
Ala Asp Tyr His Arg Val Met Tyr Ala Ser Ser Gly Ile Gly Pro Glu Lys Ala Phe Arg
261                                                                         280
Glu Leu Ala Asp Ala Leu Leu Glu Lys Gly Asn Pro Glu Met Val Asp Val Val Thr Gly
281                                                                         300
Ala Thr Val Ser Ser Gln Ser Phe Arg Arg Leu Gly Arg Ala Leu Leu Gln Ser Ala Arg
301                                                                         320
Arg Gly Glu Lys Glu Ala Ile Ile Ser Arg *** Gly Ser Gly Pro Ser Arg Cys Gly Arg
321                                                                         340
Met His Gly Thr ***
341
```

Claims

1. A fused DNA sequence which comprises a DNA sequence of a heat-resistant protein, fused directly or indirectly to a DNA sequence coding a selected protein or peptide.

2. The sequence according to Claim 1, wherein the DNA sequence of a heat-resistant protein is a DNA sequence derived from a thermophilic bacterium.

3. The sequence according to Claim 1, wherein the DNA sequence of a heat-resistant protein is a DNA sequence derived from a highly thermophilic bacterium.

4. The sequence according to Claim 1, wherein the DNA sequence of a heat-resistant protein is a DNA sequence derived from a *Thermophilus* bacterium, a *Sulfolobus* bacterium, a *Pyrococcus* bacterium, a *Thermotoga* bacte-

rium, a *Pyrobaculum* bacterium, a *Pyrodictium* bacterium, a *Thermococcus* bacterium, a *Thermodiscus* bacterium, a *Metanothermus* bacterium or a *Metanococcus* bacterium.

5. The sequence according to Claim 1, wherein the DNA sequence of a heat-resistant protein is a DNA sequence derived from a *Pyrococcus* bacterium or a *Sulfolobus* bacterium.

6. The sequence according to any one of Claims 1 to 5, wherein the DNA sequence of a heat-resistant protein is a DNA sequence of heat-resistant ferredoxin or heat-resistant adenyl kinase.

7. The sequence according to Claim 1, wherein the DNA sequence of a heat-resistant protein is a DNA sequence of ferredoxin derived from a *Pyrococcus* bacterium or adenyl kinase derived from a *Sulfolobus* bacterium.

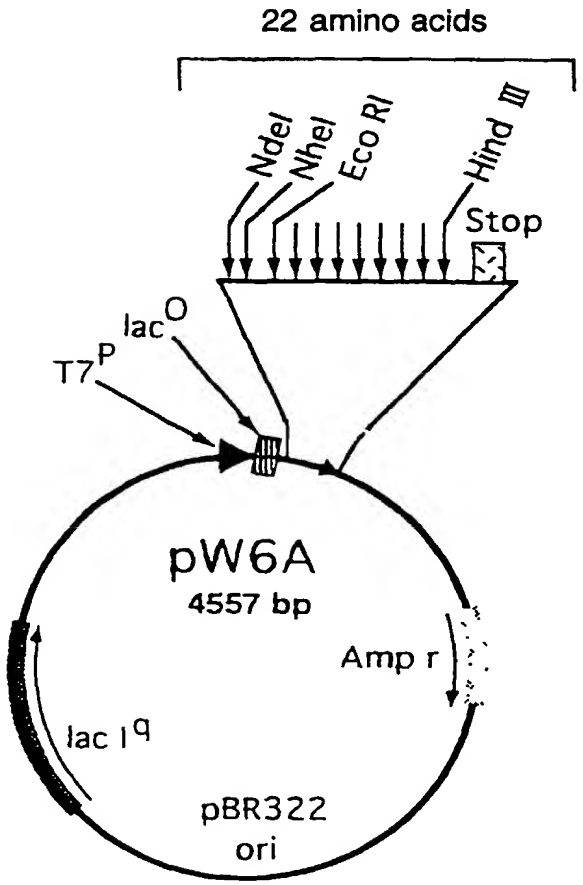
8. The sequence according to Claim 1, wherein the DNA sequence of a heat-resistant protein is a DNA sequence of ferredoxin derived from a *Pyrococcus* bacterium.

9. The sequence according to Claim 1, wherein the DNA sequence of a heat-resistant protein is a DNA sequence of adenyl kinase derived from a *Sulfolobus* bacterium.

10. A fused protein which comprises being expressed from the DNA sequence according to any one of Claims 1 to 9.

11. A method for expressing a fused protein, which comprises using the DNA sequence according to any one of Claims 1 to 9.

Fig. 1



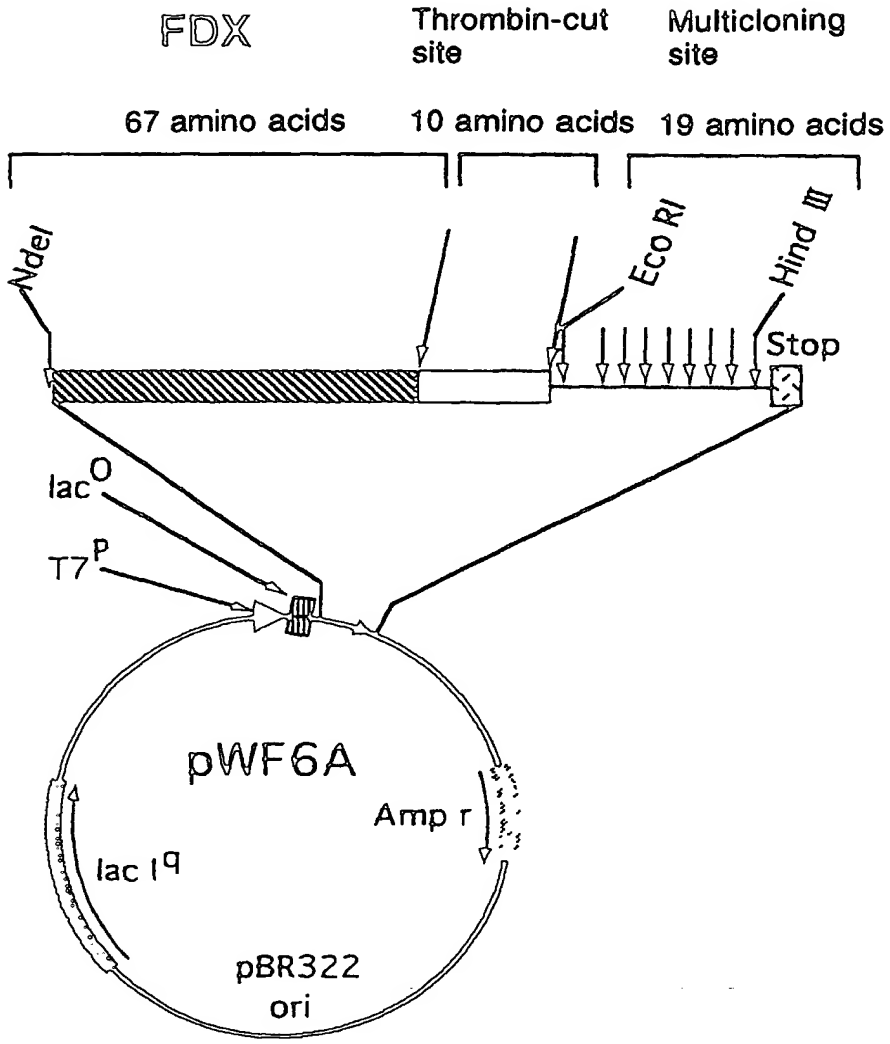
T7 P : T7 promoter

lac O : lac operator

lac I q : lac repressor

↓ : arrows of the multicloning site represent restriction enzyme sites

Fig. 2



T7 P : T7 promoter

lac O : lac operator

lac I q : lac repressor

↓ : arrows of the multicloning site represent restriction enzyme sites

Fig. 3

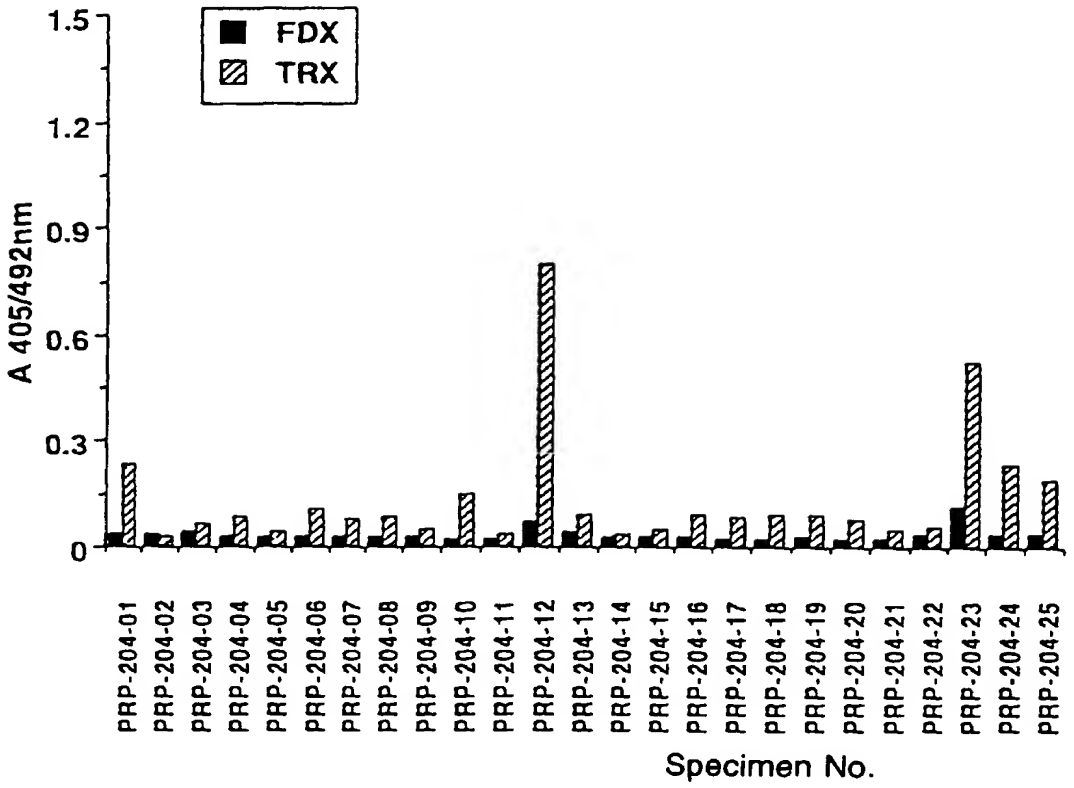


Fig. 4

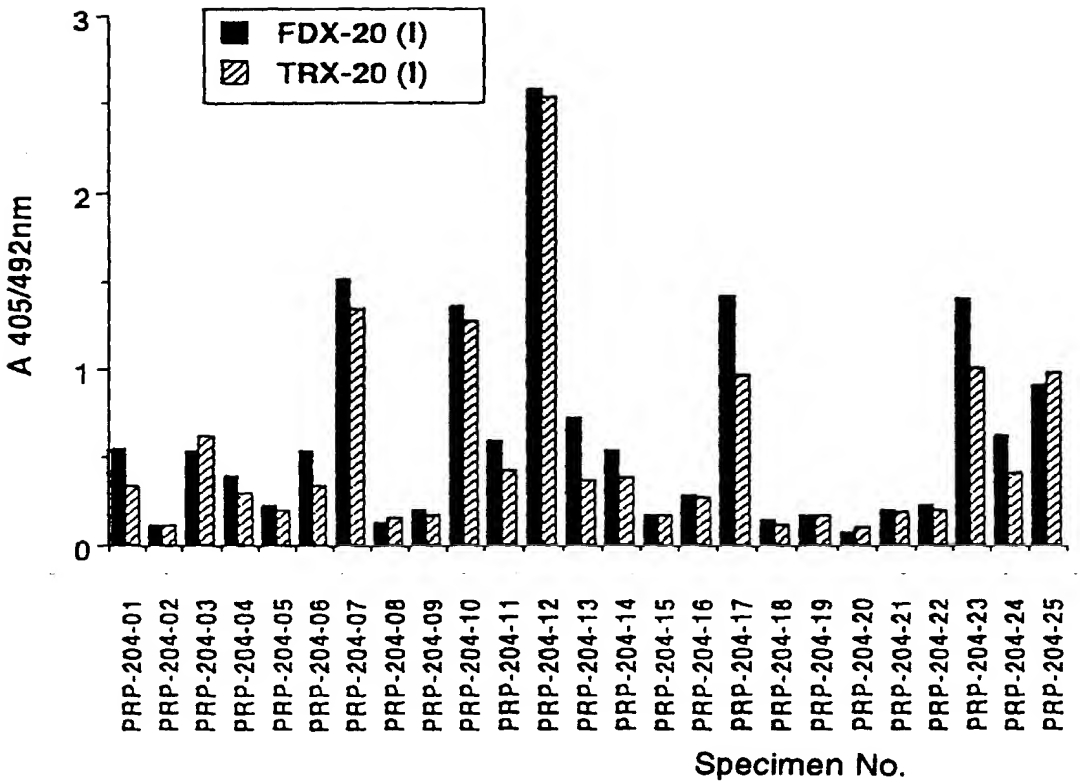


Fig. 5

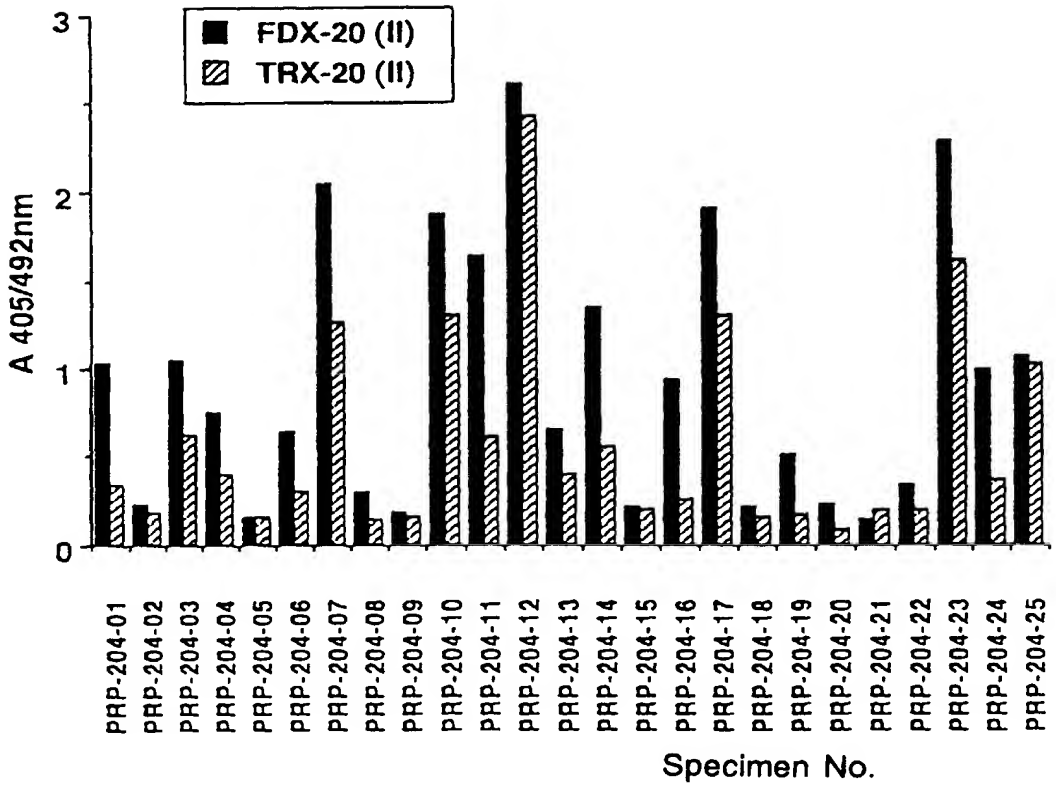


Fig. 6

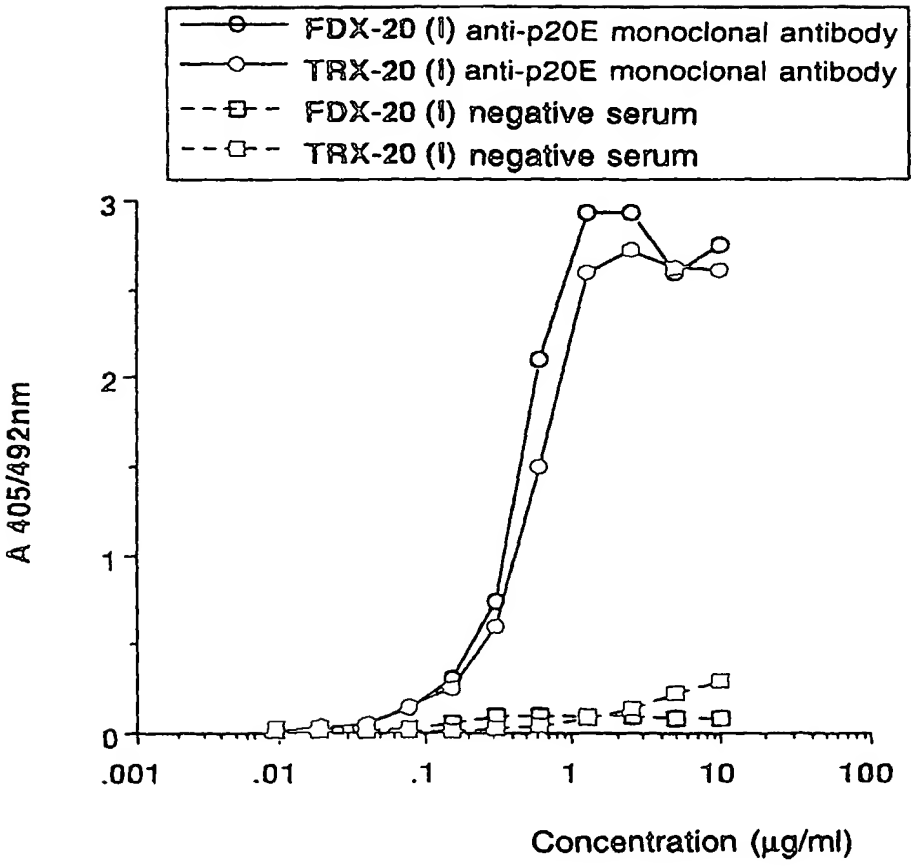


Fig. 7

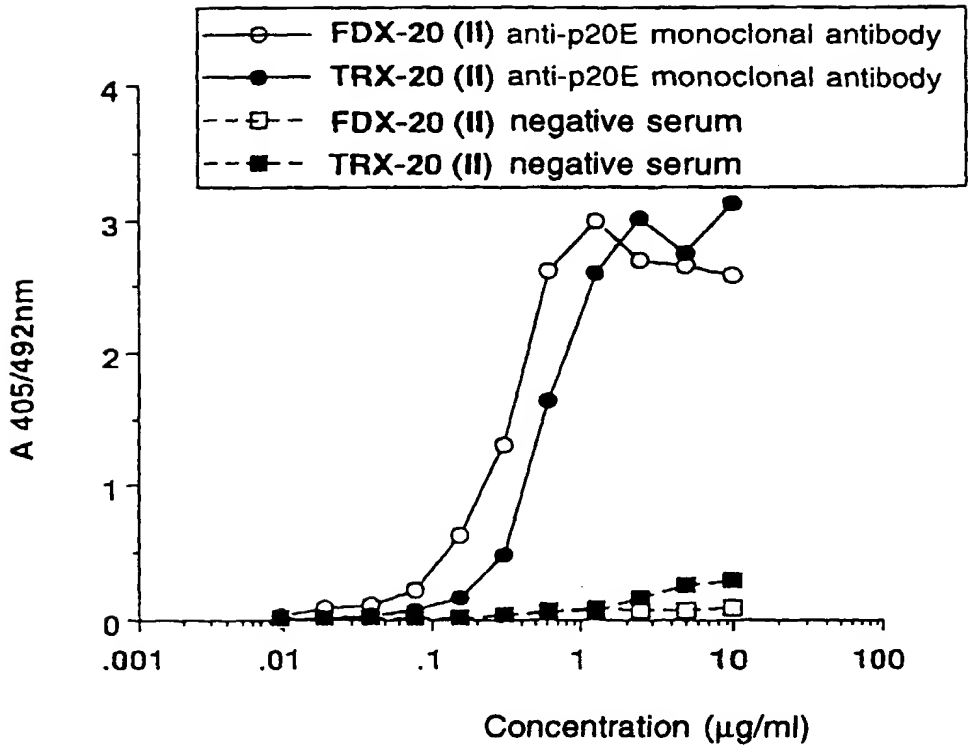


Fig. 8

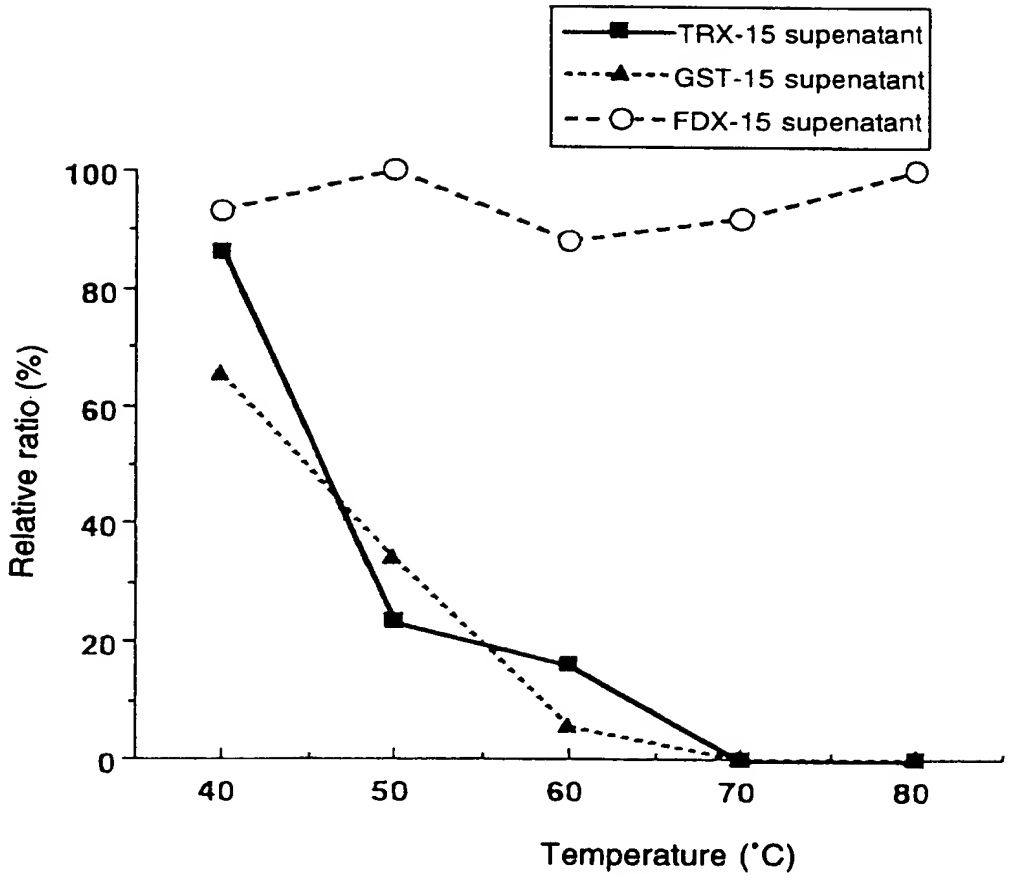


Fig. 9

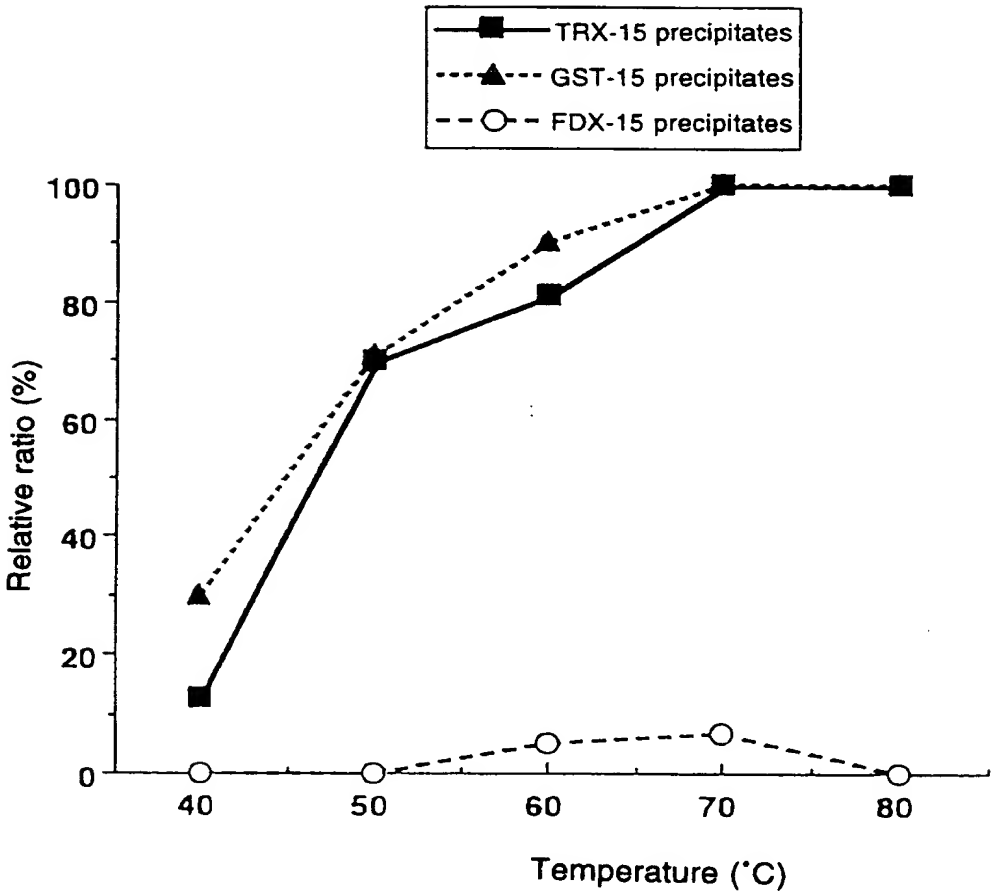


Fig. 10

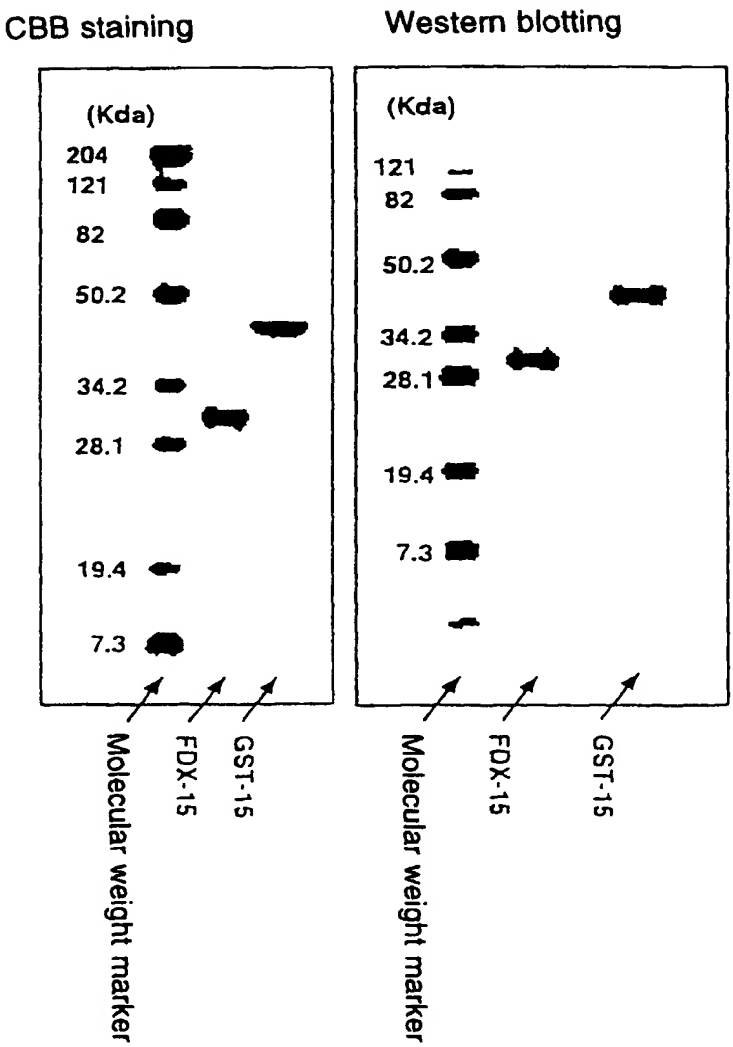
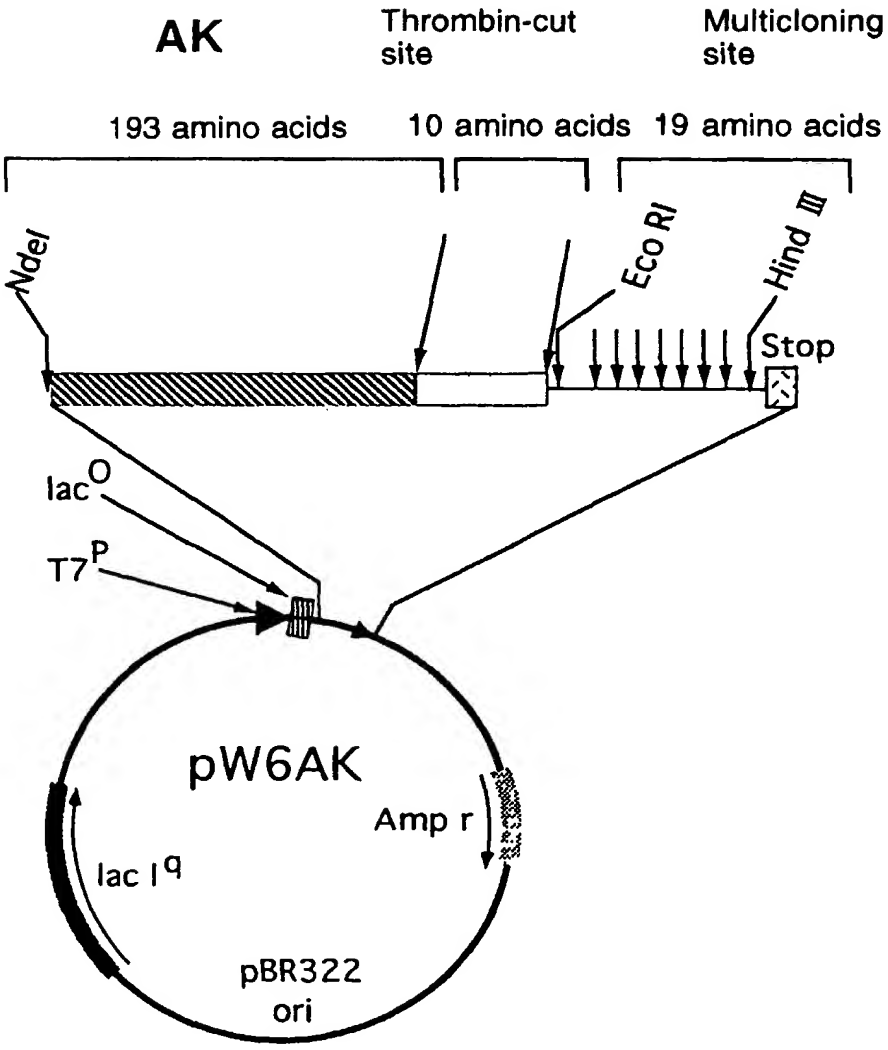


Fig. 11



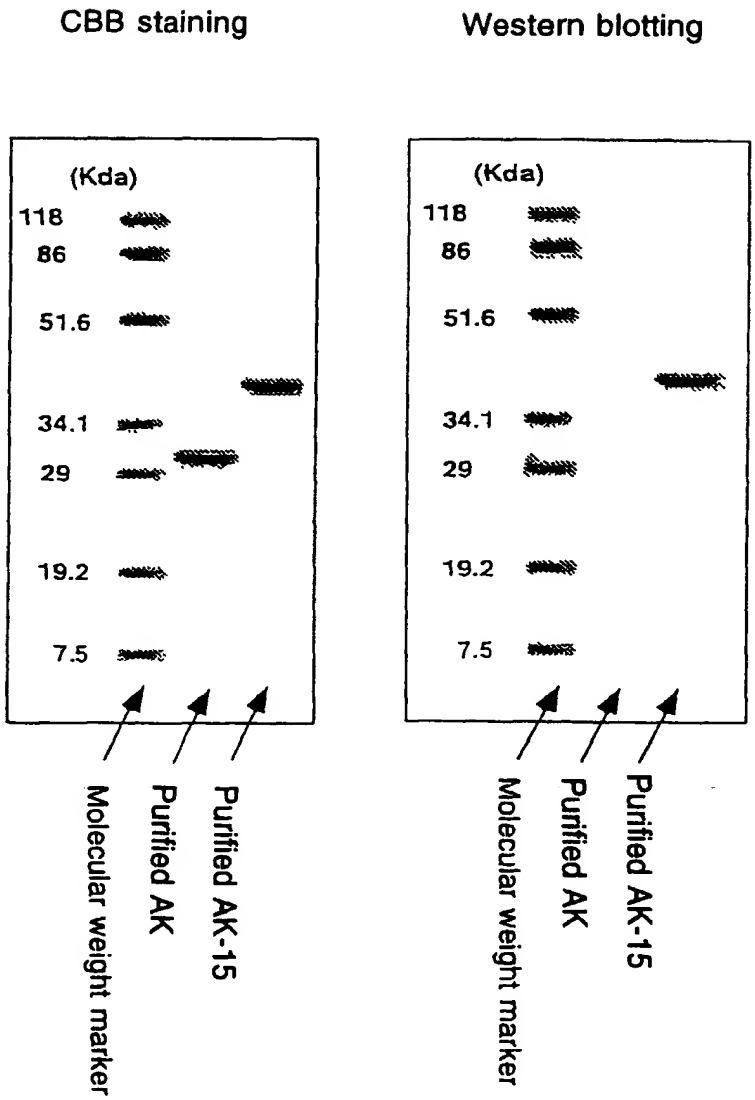
T7 p : T7 promoter

lac o : lac operator

lac I q : lac repressor

↓ : arrows of the multicloning site represent restriction enzyme sites

Fig. 12



(19) Europäisches Patentamt
Eur pean Patent Office
Office européen des brevets

(11) EP 0 781 848 A3

(12) EUROPEAN PATENT APPLICATION

(88) Date of publication A3:
01.10.1997 Bulletin 1997/40

(43) Date of publication A2:
02.07.1997 Bulletin 1997/27

(21) Application number: **96120899.8**

(22) Date of filing: **27.12.1996**

(51) Int. Cl.[6]: **C12N 15/62**, C07K 14/195, C07K 14/15, C12N 9/12, C07K 14/20, C12N 15/31, C12N 15/48, C12N 15/54

(84) Designated Contracting States:
AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE

(30) Priority: **28.12.1995 JP 352225/95**

(71) Applicant: **FUJIREBIO Inc.**
Shinjuku-ku, Tokyo 163-07 (JP)

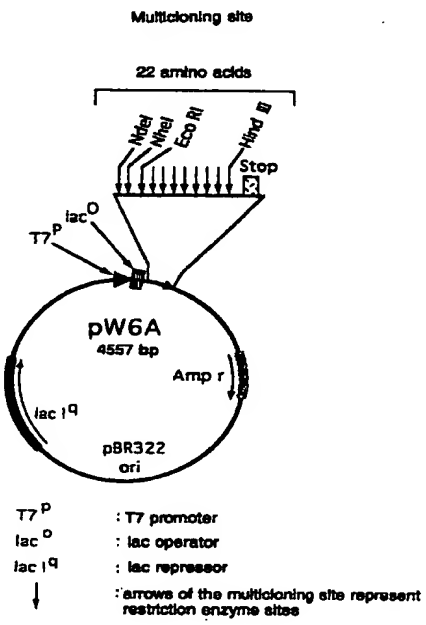
(72) Inventors:
- **Ueno, Eiichi,**
  c/o Fujirebio Inc.
  Shinjuku-ku, Tokyo (JP)
- **Fujii, Nobuyuki,**
  c/o Fujirebio Inc.
  Shinjuku-ku, Tokyo (JP)
- **Okada, Masahisa,**
  c/o Fujirebio Inc.
  Shinjuku-ku, Tokyo (JP)

(74) Representative: **Hansen, Bernd, Dr. Dipl.-Chem. et al**
Hoffmann Eitle,
Patent- und Rechtsanwälte,
Arabellastrasse 4
81925 München (DE)

(54) Fused DNA sequence, fused protein expressed from said fused DNA sequence and method for expressing said fused protein

(57) Disclosed are a fused DNA sequence which comprises a DNA sequence of a heat-resistant protein, fused directly or indirectly to a DNA sequence coding a selected protein or peptide, a fused protein expressed from the fused DNA sequence, and a method for expressing the fused protein.

Fig. 1



EP 0 781 848 A3

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 96 12 0899

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
X	PROG. BIOTECHNOL. (1995), 10(CARBOHYDRATE BIOENGINEERING), 77-84 CODEN: PBITE3;ISSN: 0921-0423, 1995, XP000647006 MORACCI, M. ET AL: "Properties and production of the.beta.-glycosidase from the thermophilic Archaeon Sulfolobus solfataricus expressed in mesophilic hosts"	1-5,10, 11	C12N15/62 C07K14/195 C07K14/15 C12N9/12 C07K14/20 C12N15/31 C12N15/48 C12N15/54
Y	* abstract * * page 80, paragraph 4 - page 83, paragraph 3 *	6-9	

D,Y	ARCHIVES OF BIOCHEMISTRY AND BIOPHYSICS, vol. 307, no. 2, December 1993, pages 405-410, XP000647072 THOMAS KATH ET AL.: "Identification, cloning, and expression of the gene for Adenylate Kinase from the thermoacidophilic Archaebacterium Sulfolobus acidocaldarius" * abstract * * page 406, right-hand column, paragraph 5 - page 408, left-hand column, paragraph 1 *	6,7,9	TECHNICAL FIELDS SEARCHED (Int.Cl.6) C12N C07K

Y	J. BACTERIOL. (1994), 176(15), 4790-3 CODEN: JOBAAY;ISSN: 0021-9193, 1994, XP000670841 HELTZEL, ANDREAS ET AL: "Cloning, expression, and molecular characterization of the gene encoding an extremely thermostable [4Fe-4S] ferredoxin from the hyperthermophilic archaeon Pyrococcus furiosus" * the whole document *	7,8	

	-/--		

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	7 August 1997	Montero Lopez, B

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)

European Patent Office

EUROPEAN SEARCH REPORT

Application Number
EP 96 12 0899

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
A	WO 87 05935 A (BATTELLE MEMORIAL INSTITUTE AT COLUMBUS) 8 October 1987 * page 16, line 6 - page 22, line 10 * -----	1,10,11	
			TECHNICAL FIELDS SEARCHED (Int.Cl.6)

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	7 August 1997	Montero Lopez, B

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P04C01)